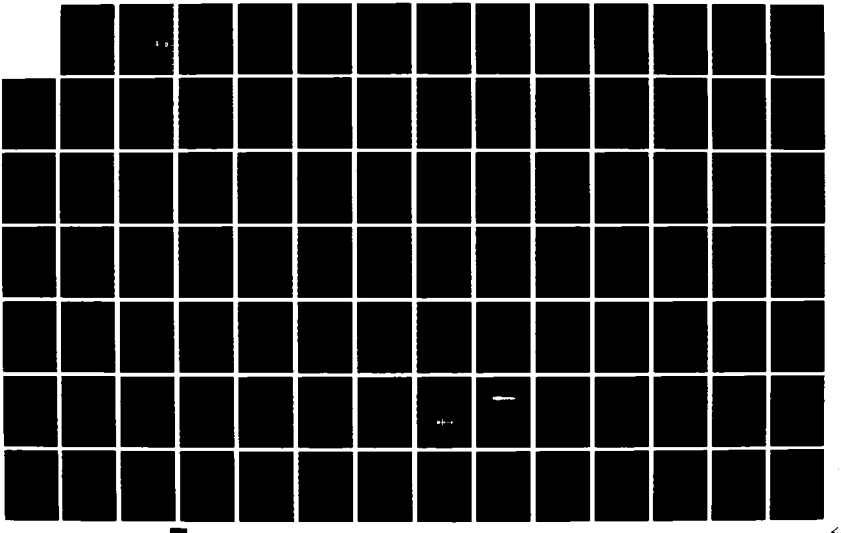
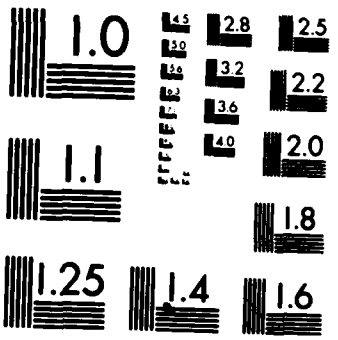
AD-A171 894 NEW ADAPTIVE IIR FILTERING ALGORITHMS(U) ILLINOIS UNIV 1/2
AT URBANA COORDINATED SCIENCE LAB H FAN AUG 86
UILU-ENG-86-2224 N00014-84-C-0149

UNCLASSIFIED F/G 12/1 NL





August 1986

UILU-ENG-86-2224

12

COORDINATED SCIENCE LABORATORY
: of Engineering

NEW ADAPTIVE IIR FILTERING ALGORITHMS

Hong Fan

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

AD-A171 094

JSEP

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS: None
2a. SECURITY CLASSIFICATION AUTHORITY: N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release, distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): UILU-ENG-86-2224	5. MONITORING ORGANIZATION REPORT NUMBER(S): N/A
6a. NAME OF PERFORMING ORGANIZATION: Coordinated Science Laboratory, Univ. of Illinois; 6b. OFFICE SYMBOL (If applicable): N/A	7a. NAME OF MONITORING ORGANIZATION: Office of Naval Research
6c. ADDRESS (City, State and ZIP Code): 1101 W. Springfield Avenue, Urbana, Illinois 61801	7b. ADDRESS (City, State and ZIP Code): 800 N. Quincy, Arlington, VA 22217
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: JSEP; 8b. OFFICE SYMBOL (If applicable): N/A	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: Contract #N00014-84-C-0149
8c. ADDRESS (City, State and ZIP Code): ONR, 800 N. Quency, Arlington, VA 22217	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
N/A	N/A	N/A	N/A

11. TITLE (Include Security Classification): New Adaptive IIR Filtering Algorithms

12. PERSONAL AUTHOR(S): Fan, Hong

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM _____ TO _____	August 1986	111

16. SUPPLEMENTARY NOTATION: N/A

17. COSATI CODES		
FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Adaptive filters, system identification, adaptive echo cancellation.

19. ABSTRACT (Continue on reverse if necessary and identify by block number): A family of adaptive IIR filtering algorithms is proposed based on the Steiglitz-McBride identification scheme. The algorithms are shown to be close approximations of one another for slow adaptation. Because of the non-vanishing gain, they are suitable for filtering applications and are simple to implement. A convergence proof is carried out using a theorem of wide-sense convergence in probability in the literature of stochastic processes. For the "sufficient order" case, the estimates can be shown to converge to the true values. While for the case of "reduced order," it is conjectured that the estimates converge to the *best fit*, which is supported by computer simulations. The major drawback is that the estimates may be biased in presence of colored disturbance. However, this does not restrict the applicability of the proposed algorithms to some important practical problems. One specific topic, adaptive echo canceling, is extensively studied and simulated for various situations. The results are favorable compared with the conventional adaptive FIR cancelers and other adaptive IIR algorithms.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
		None

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

NEW ADAPTIVE IIR FILTERING ALGORITHMS

BY

HONG FAN

Dipl., Guizhou University, 1976
M.S., University of Illinois, 1982

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1985

Urbana, Illinois

NEW ADAPTIVE IIR FILTERING ALGORITHMS

Hong Fan, Ph.D.
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, 1985

A family of adaptive IIR filtering algorithms is proposed based on the Steiglitz-McBride identification scheme. The algorithms are shown to be close approximations of one another for slow adaptation. Because of the non-vanishing gain, they are suitable for filtering applications and are simple to implement. A convergence proof is carried out using a theorem of wide-sense convergence in probability in the literature of stochastic processes. For the "sufficient order" case, the estimates can be shown to converge to the true values. While for the case of "reduced order," it is conjectured that the estimates converge to the *best fit*, which is supported by computer simulations. The major drawback is that the estimates may be biased in presence of colored disturbance. However, this does not restrict the applicability of the proposed algorithms to some important practical problems. One specific topic, adaptive echo canceling, is extensively studied and simulated for various situations. The results are favorable compared with the conventional adaptive FIR cancelers and other adaptive IIR algorithms.

ACKNOWLEDGEMENTS

I would like to thank Professor W. K. Jenkins for the initialization of my study in the adaptive filtering problem and for his continuous support during the course of this work. I would also like to thank Professor L. Praly for his helpful discussions and inspiration during his visit at the Coordinated Science Laboratory. Thanks are also due to Professor B. Hajek, Professor P. R. Kumar, and Professor D. C. Munson. I would also like to thank my wife Mary for her understanding, patience, and help. I owe her many hours of my leisure time which were used to complete this work.

TABLE OF CONTENTS

1. INTRODUCTION

The word "adapt," meaning to make fit for a new situation by modification, has appeared in filtering literature perhaps since 1960 when Widrow and Hoff published a paper on "adaptive switching circuits" [1]. Since then, an increasing stream of papers and books on adaptive filtering [2] - [42] has been appearing due to its great flexibility and vast applications. Nowadays, adaptive filters are used in prediction [7] - [10], spectral estimation [10] - [12], antenna systems [5], [12], noise cancellation [6], echo cancellation [13] - [19], channel equalization [20] - [26], time delay estimation [27] - [29], etc.. Its success is indeed tremendous.

1.1 What is Adaptive Filtering ?

"Adaptive filters" are the kind of filters which are in some sense self-designing or self-modifying. When filtering an input signal as a usual filter, an adaptive filter uses the measured input and output signals (and usually some other "reference" signal) to adjust the filter characteristics by a recursive algorithm so that the filter is constantly optimized in some statistical sense throughout the filtering process. Hence, an adaptive filter apparently performs much better than a conventional filter for the same set of conditions. This is why adaptive filters are so powerful and have drawn so much interest. Although adaptive *analog* filters do exist [2], adaptive *digital* filters are far more popular because of the rapid development in digital computers and other digital computing techniques. Throughout this dissertation, only *digital* adaptive filters will be considered and will be implied whenever "adaptive filters" are mentioned.

Two subclasses of adaptive filters can be distinguished analogous to conventional digital filters: adaptive finite impulse response (FIR) filters and adaptive infinite impulse

response (IIR) filters. Adaptive FIR filters have been the major topic in adaptive filtering literature because of the following two reasons: 1) since they have no poles, the error surfaces[1] are quadratic and thus a simple gradient search works well for the recursive adaptation algorithm; 2) also since there are no poles, the filter is always stable (provided some convergence conditions in 1) are satisfied) during the adaptation process. Note that these two reasons are distinct, although related. As mentioned earlier, the pioneering work on adaptive (FIR) filtering was done by Widrow *et al.* [1]. Widrow then went on to analyze his least mean square (LMS) algorithm [3], [4]. Meanwhile, he proposed a number of important areas of application [4] - [6]. Later, Widrow's LMS method was extended to the frequency domain where an FFT can be used for fast processing [30] - [32]. This idea was then further generalized to so-called "transform domain LMS algorithms" and was quite successful [33]. At the same time, researchers also tried to modify Widrow's LMS algorithm to achieve faster convergence and less computation [25], [34] - [36]. Finally, a lattice structure for adaptive FIR (also IIR) filters was also proposed [37] - [40]. Most of the above-mentioned progress is fairly well summarized in two recent books [41] - [42]. Indeed, the literature on adaptive FIR filtering is vast. On the contrary, the literature for adaptive IIR filtering is much less extensive. The next section explores this in more detail.

1.2 Adaptive IIR Filtering

The advantages of infinite impulse response (IIR) filters over finite impulse response (FIR) filters are well known, e.g., for the same performance, IIR filters require much less computation than FIR filters, and IIR filters can usually match physical systems well, whereas FIR filters often give only rough approximations of them. This

[1] By "error surface," it is meant the mean squared value of the output error (see Figure 1) as a function of adaptive coefficients. It is also called "performance surface."

is also true in the field of adaptive filtering. However, unlike adaptive FIR filtering, the error surfaces for adaptive IIR filters may not be unimodal, and the poles may move outside the unit circle during adaptation. These features make the adaptive IIR filtering problem much more difficult, and thus it was abandoned for a long time [43].

In 1976, Feintuch published a paper on an adaptive recursive LMS filter [44] which triggered a rebuttal [43] - [46] as well as new interest in adaptive IIR filtering. Some attempts have been made to study the behavior of error surfaces for the "system identification mode" (Figure 1) of adaptive IIR filtering. Stearns [47], based on his numerical experiments, came up with a conjecture that if the adaptive filter is of "sufficient order" (i.e., if $\hat{n}_a \geq n_a$ and $\hat{n}_b \geq n_b$), and if the driving input $x(n)$ is white noise, then the error surface $E\{e^2(n)\}$ is unimodal, where $E\{\cdot\}$ denotes the expectation. Based on his conjecture, we can roughly classify the error surfaces of the system

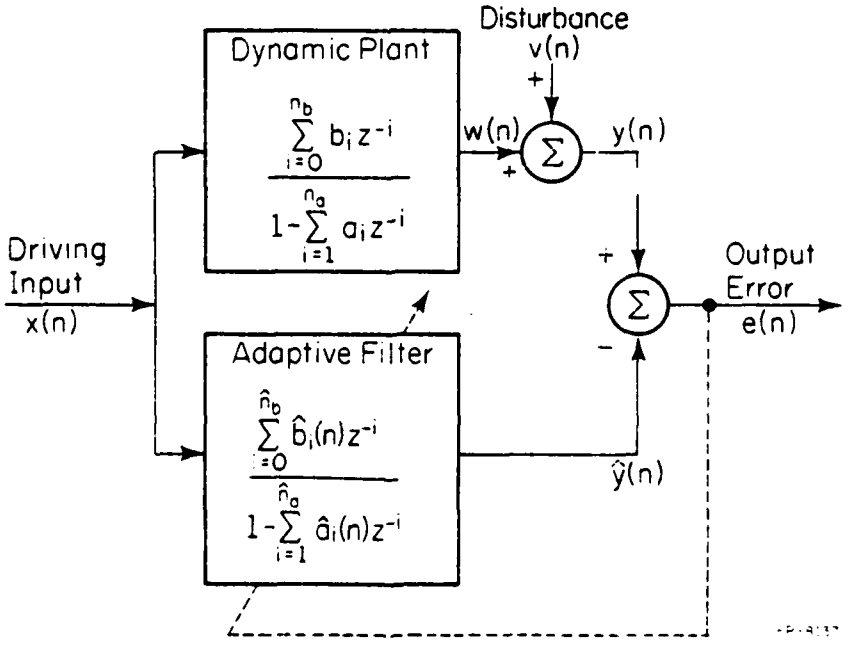


Figure 1 System identification mode of adaptive IIR filtering

identification mode in a stationary stochastic setting, with both increasing complexity and practicality:

1) sufficient order with white noise input

2) sufficient order with colored (noise) input[2]

3) reduced order with white noise input

4) reduced order with colored (noise) input

where "reduced order" refers to the situation when the conditions for the sufficient order are not satisfied. The error surfaces for 1) are, based on Stearns' conjecture, unimodal (yet to be proven). The other three classes may be multimodal. Evidence of multimodality for case 3) was shown in [45] and [47], and for case 2) was shown in [48]. A multimodal example of case 4) can be obtained by generalizing the example in [48], as illustrated in Section 2.4. Note that the proposition that the error surfaces of cases 2), 3) and 4) may be multimodal does not rule out the possibility that they may be unimodal (even for $\hat{n}_a > 0$). An example of a unimodal error surface for case 3) is found in [46]. In general, the knowledge about these error surfaces is quite limited. For error surfaces involving filter orders higher than two, the expressions can become very complicated. It also becomes difficult to display these surfaces since they are functions of more than two variables.

Early work in adaptive IIR filtering was mainly restricted to extending Widrow's LMS method of adaptive FIR filtering based on gradient search [43] - [46], [49], [50]. Typically, the technique proposed in [49] and [50] can be called "Stearns' algorithm." As just discussed, however, these algorithms may have a guaranteed global convergence only for case 1). This severely limits their usefulness. It should be pointed out here

[2] It is understood that to ensure possible parameter convergence the input should always be "persistently exciting" [52], [56] - [57].

that the direct extension of Widrow's LMS method to the IIR case results in the equation error approach [51], which shall be called LMSEE (least mean square equation error). Although LMSEE may be applicable for cases 1) and 2), it is well known that for non-zero measurement noise $v(n)$ (see Figure 1) LMSEE results in biased estimates [51] - [54].

Recently it has been realized [55] that adaptive IIR filtering, as a more general setting than the FIR, is closely related to the problem of recursive parameter estimation, a well-developed subject [56], [57]. This broadens the view of researchers in signal processing and suggests more alternatives for adaptive IIR filtering. A good example is given in [58] and [9] where the author adopted a recursive maximum likelihood estimation algorithm (RML2) for adaptive IIR filtering applications. Although RML2 is very successful in system identification applications, its direct use in adaptive filtering may not be very desirable due to its computational complexity, vanishing gain, and potential false local convergence [57] in reduced order cases 3) and 4).

Another application of recursive identification techniques to adaptive IIR filtering is represented by a family of algorithms based on the concept of hyperstability [59] - [66]. Among these, HARF (hyperstable adaptive recursive filter) was proven asymptotically convergent under the "strict positive reality" (SPR) assumption [60]. Note that HARF is potentially applicable to cases 1) and 2) while a counterexample for case 3) was reported in [62]. However, the SPR requirement is a major obstacle in the practical application of HARF [62] - [66].

Reference [67] gives a review on adaptive IIR filtering and its applications.

1.3 The Scope

Twenty years ago, Steiglitz and McBride proposed a scheme for system identification [68]. Later, Stoica and Soderstrom analyzed this scheme and gave an "on-line" version of it [69], [70]. In this dissertation, a family of stochastic approximation variants of the Steiglitz-McBride scheme is developed which shows a close relationship with LMSEE. Aiming directly at filtering applications, the contributions of this work are: i) the proposed algorithms have a non-vanishing gain which enables the algorithms to remain active and to track time-varying systems; ii) the algorithms are "on-line" and are computationally simple; and iii) an interesting phenomenon - global convergence regardless of local minima is observed which makes the algorithms promising for more general error surface cases 3) and 4). The asymptotic behavior of the proposed algorithms is similar to that of the Steiglitz-McBride scheme as analyzed in [69] and [70]. The major drawback is that the estimates are biased in the presence of colored measurement noise (disturbance). However, the global convergence phenomenon motivates the study of this family of algorithms. Note that this phenomenon has not been observed in any other adaptive IIR filtering algorithms (see, e.g., [45], [62] and [64]).

Despite the similarity in asymptotic behavior between the proposed algorithms and the Steiglitz-McBride algorithm, the convergence mechanism of the proposed algorithms is quite different from the latter because of the non-vanishing gain. Due to the lack of appropriate mathematical tools, this most practical category of adaptive IIR filtering problem, non-vanishing gain in a stochastic environment, has not received significant study [67]. Only very recently, a series of papers, e.g. [71] - [74], on "weak convergence" of adaptive algorithms for constant gains has provided us with such a tool. In this dissertation, we will use the results of Benveniste *et al.* [71] to prove

convergence of the proposed algorithm since the conditions in [71] are more "tractable" than those in others. Specifically, Benveniste's results will be used to translate the original problem into the study of an associated ordinary differential equation (ODE) regardless of the orders. The study of the ODE for sufficient order cases 1) and 2) then results in parameter convergence and output error convergence. This material is all covered in Section 2.3 (also see [75]) following the development of the algorithms in Section 2.2 (see [76], [77]). The results are verified by computer simulations in Section 2.4. For the reduced order cases 3) and 4), the ODE is much more difficult to analyze. However, computer simulations are presented in Section 2.4 which suggest global convergence for these cases. A discussion on convergence rate and other related issues is contained in Section 2.5. Chapter 3 is devoted to the application in echo cancellation. A number of computer simulations are presented for various situations. Finally, Chapter 4 concludes the dissertation by discussing some remaining issues and open questions in adaptive IIR filtering.

2. THREE NEW ADAPTIVE IIR FILTERING ALGORITHMS

2.1 Problem Formulation

The system identification mode of adaptive IIR filtering shown in Figure 1 is described by:

$$w(n)=\sum_{i=1}^{n_a} a_i\, w(n-i)+\sum_{j=0}^{n_b} b_j\, x(n-j) \tag{1a}$$

$$y(n)=w(n)+v(n) \tag{1b}$$

$$\hat{y}(n)=\sum_{i=1}^{\hat{n}_a} \hat{a}_i(n)\hat{y}(n-i)+\sum_{j=0}^{\hat{n}_b} \hat{b}_j(n)x(n-j) \tag{1c}$$

$$e(n)=y(n)-\hat{y}(n), \tag{1d}$$

where it is assumed that all the poles of the dynamic plant are inside of the unit circle, and the numerator and the denominator of its transfer function are relatively prime. If all $\hat{a}_i(n)$'s are forced to be zero, i.e., no poles are involved in the adaptive filter, we would have an adaptive FIR filter. In this case it is easily seen that the mean square error is quadratic in the $\hat{b}_j(n)$'s:

$$E\{e^2(n)\}=E\{[y(n)-\sum_{j=0}^{\hat{n}_b} \hat{b}_j(n)x(n-j)]^2\} \tag{2}$$

so that the gradient is linear in $\hat{b}_j(n)$'s:

$$\nabla_{\hat{b}_j(n)}\{E\{e^2(n)\}\}=\frac{\partial}{\partial b_j(n)}E\{e^2(n)\}=-2E\{e(n)x(n-j)\}. \tag{3}$$

Widrow's LMS algorithm is then

$$\hat{b}_j(n+1)=\hat{b}_j(n)-\tau'\nabla_{\hat{b}_j(n)}\{E\{e^2(n)\}\}=\hat{b}_j(n)+2\tau'E\{e(n)x(n-j)\}$$

$$\approx \hat{b}_j(n) + \tau e(n) x(n-j). \tag{4}$$

It was proved that the filter weights $\hat{b}_j(n)$ will converge to their optimal Wiener solution [3], [4]. However, due to the structure of the filter, this optimal solution may still yield a significant mean square error if the number of weights is not large enough. This motivates the study of the adaptive IIR filtering, i.e., allowing the freedom in adjusting $\hat{a}_i(n)$'s.

The problem of adaptive IIR filtering is not as easy as that of FIR. The difficulty arises from the recursions in $\hat{y}$. Some existing algorithms mentioned in Chapter 1 are given here.

A) Stearns' (Recursive Gradient) Algorithm [46], [49] - [50], [64], [67]:

$$\hat{a}_i(n+1) = \hat{a}_i(n) + \tau e(n) \hat{y}_i'(n); \qquad i = 1, 2, \cdots, \hat{n}_a \tag{5a}$$

$$\hat{b}_j(n+1) = \hat{b}_j(n) + \tau e(n) x_j'(n); \qquad j = 0, 1, \cdots, \hat{n}_b \tag{5b}$$

where

$$\hat{y}_i'(n) = \hat{y}(n-i) + \sum_{l=1}^{\hat{n}_a} \hat{a}_l(n) \hat{y}_i'(n-l); \qquad i = 1, 2, \cdots, \hat{n}_a \tag{5c}$$

$$x_j'(n) = x(n-j) + \sum_{l=1}^{\hat{n}_a} \hat{a}_l(n) x_j'(n-l); \qquad j = 0, 1, \cdots, \hat{n}_b. \tag{5d}$$

Note that the quantities corresponding to the gradient in (4), i.e., $\hat{y}_i'(n)$ and $x_j'(n)$ are computed recursively by (5c) and (5d). As shown by the analysis in [64], [67] and by numerical examples in [46], this algorithm is indeed a gradient algorithm and converges only to a local minimum. Thus, it may only work well for the error surface case 1).

In computing $\hat{y}_i'(n)$'s and $x_j'(n)$'s as given in (5c) and (5d), $(\hat{n}_a + \hat{n}_b + 1) \times \hat{n}_a$ storage space is required. Since $\hat{y}_i'(n)$ and $x_j'(n)$ are updated for each i and j

independently, it requires $(\hat{n}_a+\hat{n}_b+1)\times\hat{n}_a$ multiplications at each step n. Johnson mentioned in [67] a simplified version of this algorithm:

$$\hat{a}_i(n+1)=\hat{a}_i(n)+\tau e(n)\hat{y}'(n-i); \qquad i=1,2,\cdots,\hat{n}_a \tag{6a}$$

$$\hat{b}_j(n+1)=\hat{b}_j(n)+\tau e(n)x'(n-j); \qquad j=0,1,\cdots,\hat{n}_b \tag{6b}$$

where

$$\hat{y}'(n)=\hat{y}(n)+\sum_{l=1}^{\hat{n}_a}\hat{a}_l(n)\hat{y}'(n-l) \tag{6c}$$

$$x'(n)=x(n)+\sum_{l=1}^{\hat{n}_a}\hat{a}_l(n)x'(n-l). \tag{6d}$$

In other words, time shifted versions of $\hat{y}'(n)$ and $x'(n)$ are used instead of independent $\hat{y}_i'(n)$'s and $x_j'(n)$'s. In comparison with (5c) and (5d), (6c) and (6d) require only $2\hat{n}_a$ storage spaces and $2\hat{n}_a$ multiplications at each step. For slow adaptation (small τ), the results using these two versions are essentially interchangeable [67]. The Stearns' algorithm can be thought of as a direct extension of Widrow's LMS algorithm in that if $\hat{n}_a=0$, then (5) and (6) simplify to (4). Another important point is that the algorithm cannot guarantee the stability of the filter during adaptation. Thus, an expensive stability monitoring device needs to be incorporated into the algorithm.

B) Simple Hyperstable Adaptive Recursive Filter (SHARF) [59] - [67]:

The algorithm was proposed as a simplified version of the hyperstable adaptive recursive filter (HARF) [60]. For $\hat{n}_a=n_a$, $\hat{n}_b=n_b$, and $v(n)\equiv 0$, the SHARF is described as follows:

$$\hat{a}_i(n+1)=\hat{a}_i(n)+\tau e'(n)\hat{y}(n-i); \qquad i=1,2,\cdots,\hat{n}_a \tag{7a}$$

$$\hat{b}_j(n+1)=\hat{b}_j(n)+\tau e'(n)x(n-j); \qquad j=0,1,\cdots,\hat{n}_b \tag{7b}$$

$$e'(n)=e(n)+\sum_{l=1}^{\hat{n}_a} c_l\, e(n-l) \tag{7c}$$

where the coefficients c_l's are chosen so that the strictly positive realness (SPR) condition

$$\mathrm{Re}\{H(z)\}>0, \quad \text{for all} \quad |z|=1; \qquad H(z)=\frac{1+\sum_{l=1}^{\hat{n}_a} c_l\, z^{-l}}{1-\sum_{l=1}^{\hat{n}_a} a_l\, z^{-l}} \tag{8}$$

is satisfied. While SHARF is capable of dealing with error surface cases 1) and 2) and has guaranteed stability, the SPR requirement is an obstacle, since some *a priori* knowledge about the unknown plant parameters a_l's is needed. After exchanging ideas through extensive correspondence in journals and conference proceedings [62] - [66], researchers came up with a modified SHARF algorithm that eliminates the SPR requirement as well as, unfortunately, the guaranteed stability [63], [66]:

$$\hat{a}_i(n+1)=\hat{a}_i(n)+\tau e'(n)\hat{y}(n-i); \qquad i=1,2,\cdots,\hat{n}_a \tag{9a}$$

$$\hat{b}_j(n+1)=\hat{b}_j(n)+\tau e'(n)x(n-j); \qquad j=0,1,\cdots,\hat{n}_b \tag{9b}$$

$$\hat{c}_k(n+1)=\hat{c}_k(n)+\tau e'(n)e(n-k); \qquad k=1,2,\cdots,\hat{n}_a \tag{9c}$$

$$e'(n)=e(n)+\sum_{l=1}^{\hat{n}_a} \hat{c}_l(n)\, e(n-l) \tag{9d}$$

provided that

$$\hat{C}(n,z^{-1})=1+\sum_{l=1}^{\hat{n}_a} \hat{c}_l(n)\, z^{-l} \quad \textit{has all roots within the unit circle} \tag{10}$$

during adaptation, which requires also a stability monitoring device [66], [67].

Note that a rigorous convergence proof applies only to HARF, which is computationally much more complicated [60]. The SPR requirement (8) and the stability requirement (10) were imposed on HARF. Nevertheless, for small τ SHARF remains a close approximation to HARF, and thus these requirements are essentially applicable to SHARF [59].

For reduced order cases 3) and 4), SHARF (or HARF) fails to converge to the desired global minimum as demonstrated in [62].

C) Least Mean Square Equation Error (LMSEE) Algorithm [51] - [54]:

Another way of directly extending Widrow's LMS algorithm to the IIR case is by filtering the output error through an all-zero (post) filter identical to the denominator of the adaptive filter, so that this filtered output error, or the "equation error," is linear in $\hat{a}_i(n)$'s and $\hat{b}_j(n)$'s [52]:

$$e'(n)=e(n)-\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)e(n-i),$$

$$=y(n)-\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)y(n-i)-[\hat{y}(n)-\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)\hat{y}(n-i)]$$

$$=y(n)-\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)y(n-i)-\sum_{j=0}^{\hat{n}_b}\hat{b}_j(n)x(n-j). \qquad (11)$$

It is then a simple matter to minimize $E\{e'^2(n)\}$ by the widely used LMS method, i.e.,

$$\frac{\partial}{\partial\hat{a}_i(n)}E\{e'^2(n)\}\approx 2e'(n)\frac{\partial}{\partial\hat{a}_i(n)}e'(n)=-2e'(n)y(n-i): \quad i=1,2,\cdots,\hat{n}_a \qquad (12a)$$

$$\frac{\partial}{\partial\hat{b}_j(n)}E\{e'^2(n)\}\approx 2e'(n)\frac{\partial}{\partial\hat{b}_j(n)}e'(n)=-2e'(n)x(n-j): \quad j=0,1,\cdots,\hat{n}_b \qquad (12b)$$

such that

$$\hat{a}_i(n+1)=\hat{a}_i(n)+\tau e'(n)y(n-i); \quad i=1,2,\cdots,\hat{n}_a \tag{13a}$$

$$\hat{b}_j(n+1)=\hat{b}_j(n)+\tau e'(n)x(n-j); \quad j=0,1,\cdots,\hat{n}_b. \tag{13b}$$

This approach minimizes the (squared) equation error but does not necessarily minimize the (squared) output error unless the minimum values are zero (one being zero implies the other as clearly seen in (11), assuming $1-\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)z^{-i}$ having all its roots within the unit circle), which corresponds to the error surface cases 1) and 2) when the disturbance $v(n)\equiv 0$. For non-zero $v(n)$ the parameter estimates are generally biased [51] - [54]. Debiasing requires *a priori* knowledge about the statistics of $v(n)$ and thus may not be desirable [51].

Other adaptive IIR filtering algorithms also exist, e.g., RML2 adopted by Friedlander [58]. However, as mentioned before, due to the computational complexity and other reasons, they are not presented in this dissertation.

2.2 Development of the New Algorithms

The family of the new algorithms is developed based on the LMSEE algorithm. Consider for the moment $x'(n)$ as the input, $e'(n)$ as the output error, and $v'(n)$ as the disturbance in Figure 2 (this subsystem is the same as Figure 1), $e(n)$ then becomes the "equation error" as described in Section 2.1 and can be expressed as

$$e(n)=e'(n)-\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)e'(n-i), \tag{14}$$

where

$$e'(n)=y'(n)-\hat{y}'(n) \tag{15a}$$

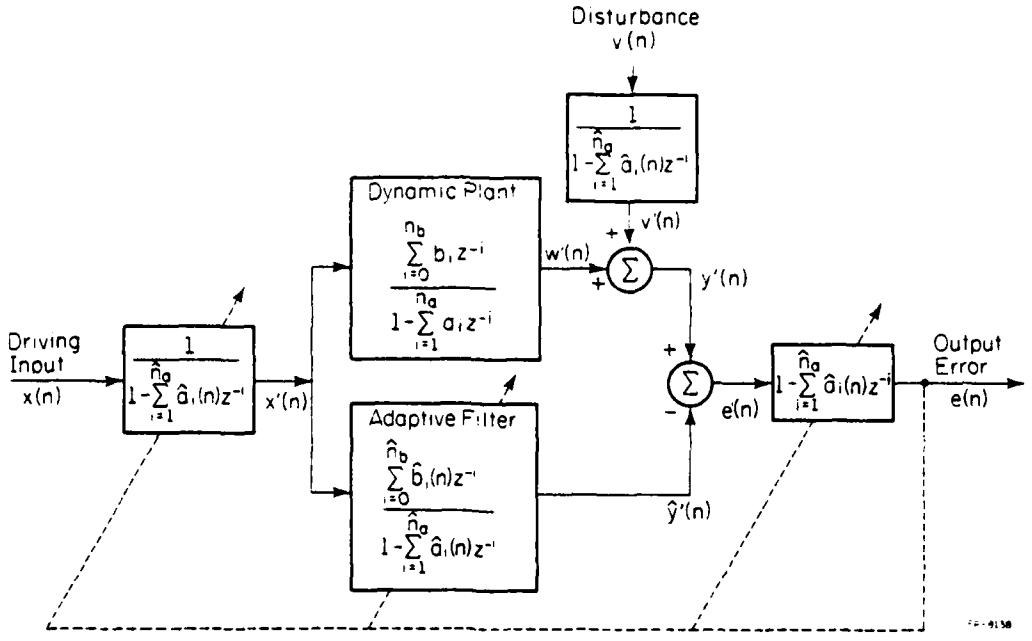


Figure 2 System identification mode of the proposed algorithm

$$w'(n)=\sum_{i=1}^{n_a} a_i\, w'(n-i)+\sum_{j=0}^{n_b} b_j\, x'(n-j) \tag{15b}$$

$$y'(n)=w'(n)+v'(n) \tag{15c}$$

$$v'(n)=v(n)+\sum_{i=1}^{\hat{n}_a} \hat{a}_i(n)\, v'(n-i) \tag{15d}$$

$$\hat{y}'(n)=\sum_{i=1}^{\hat{n}_a} \hat{a}_i(n)\, \hat{y}'(n-i)+\sum_{j=0}^{\hat{n}_b} \hat{b}_j(n)\, x'(n-j), \tag{15e}$$

analogous to (1). Substituting (15a) and (15e) into (14) results in

$$e(n)=y'(n)-\sum_{i=1}^{\hat{n}_a} \hat{a}_i(n)\, y'(n-i)-\left[\hat{y}'(n)-\sum_{i=1}^{\hat{n}_a} \hat{a}_i(n)\, \hat{y}'(n-i)\right]$$

$$=y'(n)-\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)y'(n-i)-\sum_{j=0}^{\hat{n}_b}\hat{b}_j(n)x'(n-j). \tag{16}$$

Here $e(n)$ is linear in $\hat{a}_i(n)$'s and $\hat{b}_i(n)$'s assuming y' and x' are independent of these adaptive coefficients for the moment. Then minimization of $E\{e^2(n)\}$ yields

$$\frac{\partial}{\partial\hat{a}_i(n)}E\{e^2(n)\}\approx 2e(n)\frac{\partial}{\partial\hat{a}_i(n)}e(n)=-2e(n)y'(n-i); \quad i=1,2,\cdots,\hat{n}_a \tag{17a}$$

$$\frac{\partial}{\partial\hat{b}_j(n)}E\{e^2(n)\}\approx 2e(n)\frac{\partial}{\partial\hat{b}_j(n)}e(n)=-2e(n)x'(n-j); \quad j=0,1,\cdots,\hat{n}_b \tag{17b}$$

such that

$$\hat{a}_i(n+1)=\hat{a}_i(n)+\tau e(n)y'(n-i); \quad i=1,2,\cdots,\hat{n}_a \tag{18a}$$

$$\hat{b}_j(n+1)=\hat{b}_j(n)+\tau e(n)x'(n-j); \quad j=0,1,\cdots,\hat{n}_b. \tag{18b}$$

As stated before, this approach minimizes the (squared) equation error but not the (squared) output error. To circumvent this, an all-pole filter which is the inverse of the post-filter is added as a pre-filter at the input, as shown in Figure 2. Note that the overall response from $x(n)$ to $e(n)$ is not the same as that of Figure 1, due to the time-varying nature of the filters. However, for slow adaptation it is a close approximation to that of Figure 1. Moreover, at a convergence point where all adaptive coefficients tend to be constant, these two systems become exactly the same. Thus $e(n)$, the equation error with respect to $x'(n)$, can be thought of as the output error for the input $x(n)$. Then the quadratic minimization scheme for the equation error ((14) - (18)) can be applied provided $x'(n)$ and $y'(n)$ are independent of $\hat{a}_i(n)$'s and $\hat{b}_i(n)$'s. Due to the pre-filter, this is apparently not true. However, we observe that Equation (17b) is valid, since $x'(n)$ and $y'(n)$ are independent of $\hat{b}_i(n)$'s, and the derivative of $e(n)$ with respect to $\hat{a}_i(n)$, considering (15b), (15c), (15d) and (16), is

$$\frac{\partial}{\partial \hat{a}_i(n)} e(n) = \frac{\partial}{\partial \hat{a}_i(n)} [w'(n) + v'(n)] - \sum_{l=1}^{\hat{n}_a} \hat{a}_l(n) \frac{\partial}{\partial \hat{a}_i(n)} y'(n-l) - y'(n-i)$$

$$- \sum_{j=0}^{\hat{n}_b} \hat{b}_j(n) \frac{\partial}{\partial \hat{a}_i(n)} x'(n-j)$$

$$= \sum_{l=1}^{n_a} a_l \frac{\partial}{\partial \hat{a}_i(n)} w'(n-l) + \sum_{j=0}^{n_b} b_j \frac{\partial}{\partial \hat{a}_i(n)} x'(n-j) + \sum_{l=1}^{\hat{n}_a} \hat{a}_l(n) \frac{\partial}{\partial \hat{a}_i(n)} v'(n-l)$$

$$+ v'(n-i) - \sum_{l=1}^{\hat{n}_a} \hat{a}_l(n) \frac{\partial}{\partial \hat{a}_i(n)} [w'(n-l) + v'(n-l)] - \sum_{j=0}^{\hat{n}_b} \hat{b}_j(n) \frac{\partial}{\partial \hat{a}_i(n)} x'(n-j)$$

$$- y'(n-i)$$

$$= \sum_{l=1}^{n_1} [a_l - \hat{a}_l(n)] \frac{\partial}{\partial \hat{a}_i(n)} w'(n-l) + \sum_{j=0}^{n_2} [b_j - \hat{b}_j(n)] \frac{\partial}{\partial \hat{a}_i(n)} x'(n-j) - w'(n-i)$$

$$\approx -w'(n-i); \qquad i = 1, 2, \cdots, \hat{n}_a. \tag{19}$$

where $n_1 = \max\{n_a, \hat{n}_a\}$ and $n_2 = \max\{n_b, \hat{n}_b\}$. From this it is seen that Equation (17a) is an approximation in that w' is replaced by y' and the two summation terms are ignored in the last step of (19). Note that this is a reasonable approximation for the following reasons. First, because $v(n)$ [and hence $v'(n)$] is unmeasurable, $y'(n-j)$ is the only measurable quantity that gives an approximation for $w'(n-j)$. For $v(n) \equiv 0$ this approximation becomes exact. Second, for the sufficient order case $\hat{n}_a = n_a$, $\hat{n}_b = n_b$, the two summation terms in (19) indeed approach zero near the (parameter) convergence point and hence resulting in a gradient nature at that point which guarantees the local convergence at that (global minimum) point. Whereas at other points (possibly other local minima) the two ignored summation terms are not zero, and hence the algorithm does not exhibit a gradient behavior, which may prevent itself

$$=y'(n)-\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)y'(n-i)-\sum_{j=0}^{\hat{n}_b}\hat{b}_j(n)x'(n-j). \qquad (16)$$

Here $e(n)$ is linear in $\hat{a}_i(n)$'s and $\hat{b}_i(n)$'s assuming y' and x' are independent of these adaptive coefficients for the moment. Then minimization of $E\{e^2(n)\}$ yields

$$\frac{\partial}{\partial\hat{a}_i(n)}E\{e^2(n)\}\approx 2e(n)\frac{\partial}{\partial\hat{a}_i(n)}e(n)=-2e(n)y'(n-i); \quad i=1,2,\cdots,\hat{n}_a \qquad (17a)$$

$$\frac{\partial}{\partial\hat{b}_j(n)}E\{e^2(n)\}\approx 2e(n)\frac{\partial}{\partial\hat{b}_j(n)}e(n)=-2e(n)x'(n-j); \quad j=0,1,\cdots,\hat{n}_b \qquad (17b)$$

such that

$$\hat{a}_i(n+1)=\hat{a}_i(n)+\tau e(n)y'(n-i); \quad i=1,2,\cdots,\hat{n}_a \qquad (18a)$$

$$\hat{b}_j(n+1)=\hat{b}_j(n)+\tau e(n)x'(n-j); \quad j=0,1,\cdots,\hat{n}_b. \qquad (18b)$$

As stated before, this approach minimizes the (squared) equation error but not the (squared) output error. To circumvent this, an all-pole filter which is the inverse of the post-filter is added as a pre-filter at the input, as shown in Figure 2. Note that the overall response from $x(n)$ to $e(n)$ is not the same as that of Figure 1, due to the time-varying nature of the filters. However, for slow adaptation it is a close approximation to that of Figure 1. Moreover, at a convergence point where all adaptive coefficients tend to be constant, these two systems become exactly the same. Thus $e(n)$, the equation error with respect to $x'(n)$, can be thought of as the output error for the input $x(n)$. Then the quadratic minimization scheme for the equation error ((14) - (18)) can be applied provided $x'(n)$ and $y'(n)$ are independent of $\hat{a}_i(n)$'s and $\hat{b}_i(n)$'s. Due to the pre-filter, this is apparently not true. However, we observe that Equation (17b) is valid, since $x'(n)$ and $y'(n)$ are independent of $\hat{b}_i(n)$'s, and the derivative of $e(n)$ with respect to $\hat{a}_i(n)$, considering (15b), (15c), (15d) and (16), is

$$\frac{\partial}{\partial \hat{a}_i(n)} e(n) = \frac{\partial}{\partial \hat{a}_i(n)} [w'(n) + v'(n)] - \sum_{l=1}^{\hat{n}_a} \hat{a}_l(n) \frac{\partial}{\partial \hat{a}_i(n)} y'(n-l) - y'(n-i)$$

$$- \sum_{j=0}^{\hat{n}_b} \hat{b}_j(n) \frac{\partial}{\partial \hat{a}_i(n)} x'(n-j)$$

$$= \sum_{l=1}^{n_a} a_l \frac{\partial}{\partial \hat{a}_i(n)} w'(n-l) + \sum_{j=0}^{n_b} b_j \frac{\partial}{\partial \hat{a}_i(n)} x'(n-j) + \sum_{l=1}^{\hat{n}_a} \hat{a}_l(n) \frac{\partial}{\partial \hat{a}_i(n)} v'(n-l)$$

$$+ v'(n-i) - \sum_{l=1}^{\hat{n}_a} \hat{a}_l(n) \frac{\partial}{\partial \hat{a}_i(n)} [w'(n-l) + v'(n-l)] - \sum_{j=0}^{\hat{n}_b} \hat{b}_j(n) \frac{\partial}{\partial \hat{a}_i(n)} x'(n-j)$$

$$- y'(n-i)$$

$$= \sum_{l=1}^{n_1} [a_l - \hat{a}_l(n)] \frac{\partial}{\partial \hat{a}_i(n)} w'(n-l) + \sum_{j=0}^{n_2} [b_j - \hat{b}_j(n)] \frac{\partial}{\partial \hat{a}_i(n)} x'(n-j) - w'(n-i)$$

$$\approx -w'(n-i); \qquad i = 1, 2, \cdots, \hat{n}_a. \tag{19}$$

where $n_1 = \max\{n_a, \hat{n}_a\}$ and $n_2 = \max\{n_b, \hat{n}_b\}$. From this it is seen that Equation (17a) is an approximation in that w' is replaced by y' and the two summation terms are ignored in the last step of (19). Note that this is a reasonable approximation for the following reasons. First, because $v(n)$ [and hence $v'(n)$] is unmeasurable, $y'(n-j)$ is the only measurable quantity that gives an approximation for $w'(n-j)$. For $v(n) \equiv 0$ this approximation becomes exact. Second, for the sufficient order case $\hat{n}_a = n_a$, $\hat{n}_b = n_b$, the two summation terms in (19) indeed approach zero near the (parameter) convergence point and hence resulting in a gradient nature at that point which guarantees the local convergence at that (global minimum) point. Whereas at other points (possibly other local minima) the two ignored summation terms are not zero, and hence the algorithm does not exhibit a gradient behavior, which may prevent itself

from being trapped at a local minimum as other algorithms are. For reduced order cases the situation is more complicated and it may also very well be these ignored terms that make the algorithm converge globally (see Section 2.4). Another more heuristic argument for this approximation may be as follows. At each step of adaptation, the input $x(n)$ is pre-filtered, producing $x'(n)$, before the adaptive coefficients are updated. It is thus assumed that the pre-filter is a constant filter with coefficients fixed at $\hat{a}_i(n)$'s at each step of adaptation n. Then $x'(n)$ and $y'(n)$ can be thought of as independent of the coefficients $\hat{a}_i(n)$'s and $\hat{b}_i(n)$'s in the adaptive filter and the post-filter at the time instant n. The quadratic scheme (14) - (18) can still be applied. After updating, the coefficients in the pre-filter are also changed, resulting in a new fixed pre-filter for the next adaptation. This argument is more valid for the "IF" algorithm appearing next. In short, Eqns. (14), (15), and (18) together with

$$x'(n)=x(n)+\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)x'(n-i) \tag{20}$$

constitute one version (system identification mode) of the proposed algorithm. It shall be referred to as SIM algorithm in the sequel.

Note that since $v(n)$ is not measurable, (15d) is not practically realizable. Thus SIM algorithm is applicable only when $v(n)\equiv 0$. Also, in the general adaptive filtering mode, $y(n)$ is usually the given data ("desired response") so that $y'(n)$ cannot be obtained by feeding $x'(n)$ into the (unknown) dynamic plant and measuring the output, as described by (15b) - (15d). In this case, another version of the proposed algorithm can be comprised of Equations (14), (15a), (15e), (18), (20) and

$$y'(n)=y(n)+\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)y'(n-i). \tag{21}$$

where $y(n)$ is the given data which can, for example, be modeled by (1a) and (1b).

Note that the two pre-filters ((20) and (21)) are necessary to obtain $x'(n)$ and $y'(n)$ to update the coefficients. However, it is not necessary to calculate $e(n)$ using (14), (15a), (15e), (20), and (21). In fact, the pre-filter and the post-filter approximately cancel as argued before and thus $e(n)$ can be calculated directly from (1d). Therefore, the adaptive filtering mode of the proposed algorithm (AFM algorithm) is described by Equations (1c), (1d), (18), (20), and (21). Figure 3 shows the implementation. Note that besides the adaptive filter and the coefficient updating mechanism (18), only two additional filters are used. Thus, it is computationally simple and efficient.

Analogous to (5) and (6), as in Stearns' algorithm, a somewhat more complicated version of the proposed algorithm can be obtained as follows. Instead of using a time-shifted version of y' and x', i.e., $y'(n-i)$ and $x'(n-j)$ as in (18), a set of independently filtered quantities, $y_i'(n)$ and $x_j'(n)$ is used. They are obtained by

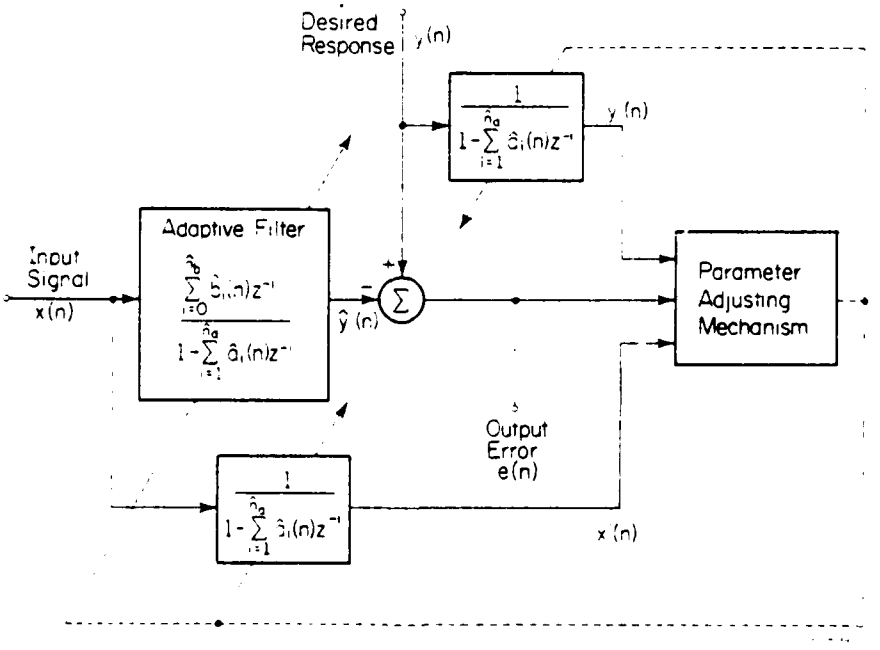


Figure 3 Adaptive filtering mode of the proposed algorithm

$$y_i'(n)=y(n-i)+\sum_{l=1}^{\hat{n}_a}\hat{a}_l(n)y_i'(n-l);\qquad i=1,2,\cdots,\hat{n}_a \tag{22a}$$

$$x_j'(n)=x(n-j)+\sum_{l=1}^{\hat{n}_a}\hat{a}_l(n)x_j'(n-l);\qquad j=0,1,\cdots,\hat{n}_b. \tag{22b}$$

This can be applied to either SIM or AFM. Since AFM algorithm is more practical and can also be used for the system identification purpose, only an AFM version of this independent filtering (IF) algorithm is considered. As in the Stearns' algorithm, the storage space and the number of multiplications at each step will increase from $2\hat{n}_a$ (AFM) to $(\hat{n}_a+\hat{n}_b+1)\times\hat{n}_a$ (IF). Although the IF algorithm does not seem to be computationally more advantageous than the others, it allows the development of a rigorous convergence proof presented in the next section, which will guide the use of the AFM algorithm, since it is a close approximation of the IF algorithm for slow adaptation. One can easily see that if $\hat{a}_i(n)$'s are not time-varying, then (22) is the same as the time-shifted versions of (20) and (21). Although theoretically these three algorithms are all different due to the time-varying nature of the filters, asymptotically they behave the same because as the filter coefficients converge, these filters become more time-invariant (see Section 2.3). Practically, for slow adaptation (usually the case) even at points other than the convergent ones, these algorithms still behave almost the same. Our computer simulations show that the convergence paths and mean square error progressions using these three algorithms are almost indistinguishable (see Section 2.4).

Finally, all the three proposed algorithms are summarized below.

A) SIM Algorithm (Figure 2):

$$\hat{a}_i(n+1)=\hat{a}_i(n)+\tau e(n)y'(n-i);\qquad i=1,2,\cdots,\hat{n}_a \tag{23a}$$

$$\hat{b}_j(n+1)=\hat{b}_j(n)+\tau e(n)x'(n-j): \qquad j=0,1,\cdots,\hat{n}_b \tag{23b}$$

$$x'(n)=x(n)+\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)x'(n-i) \tag{23c}$$

$$e(n)=e'(n)-\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)e'(n-i) \tag{23d}$$

$$e'(n)=y'(n)-\hat{y}'(n) \tag{23e}$$

with

$$w'(n)=\sum_{i=1}^{n_a}a_i w'(n-i)+\sum_{j=0}^{n_b}b_j x'(n-j) \tag{23f}$$

$$y'(n)=w'(n)+v'(n) \tag{23g}$$

$$v'(n)=v(n)+\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)v'(n-i) \tag{23h}$$

$$\hat{y}'(n)=\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)\hat{y}'(n-i)+\sum_{j=0}^{\hat{n}_b}\hat{b}_j(n)x'(n-j) \tag{23i}$$

B) AFM Algorithm (Figure 3):

$$\hat{a}_i(n+1)=\hat{a}_i(n)+\tau e(n)y'(n-i): \qquad i=1,2,\cdots,\hat{n}_a \tag{24a}$$

$$\hat{b}_j(n+1)=\hat{b}_j(n)+\tau e(n)x'(n-j): \qquad j=0,1,\cdots,\hat{n}_b \tag{24b}$$

$$y'(n)=y(n)+\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)y'(n-i) \tag{24c}$$

$$x'(n)=x(n)+\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)x'(n-i) \tag{24d}$$

$$e(n)=y(n)-\hat{y}(n) \tag{24e}$$

with

$$\hat{y}(n)=\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)\hat{y}(n-i)+\sum_{j=0}^{\hat{n}_b}\hat{b}_j(n)x(n-j) \tag{24f}$$

C) IF Algorithm:

$$\hat{a}_i(n+1)=\hat{a}_i(n)+\tau e(n)y_i'(n): \qquad i=1,2,\cdots,\hat{n}_a \tag{25a}$$

$$\hat{b}_j(n+1)=\hat{b}_j(n)+\tau e(n)x_j'(n): \qquad j=0,1,\cdots,\hat{n}_b \tag{25b}$$

$$y_i'(n)=y(n-i)+\sum_{l=1}^{\hat{n}_a}\hat{a}_l(n)y_i'(n-l): \qquad i=1,2,\cdots,\hat{n}_a \tag{25c}$$

$$x_j'(n)=x(n-j)+\sum_{l=1}^{\hat{n}_a}\hat{a}_l(n)x_j'(n-l): \qquad j=0,1,\cdots,\hat{n}_b \tag{25d}$$

$$e(n)=y(n)-\hat{y}(n) \tag{25e}$$

with

$$\hat{y}(n)=\sum_{i=1}^{\hat{n}_a}\hat{a}_i(n)\hat{y}(n-i)+\sum_{j=0}^{\hat{n}_b}\hat{b}_j(n)x(n-j) \tag{25f}$$

It is interesting to note that the only difference between AFM. IF algorithms and the corresponding versions of Stearns' algorithm is that in these algorithms y is filtered to obtain y' whereas in Stearns' algorithm $\hat{y}$ is filtered to obtain $\hat{y}'$.

2.3 A Convergence Proof

The convergence proof presented in this section is for the IF algorithm only. A convergence proof for the AFM algorithm has not been developed yet. However, as

mentioned before, the AFM algorithm is a close approximation of the IF algorithm for slow adaptation. Thus, the theory developed for the IF algorithm can also be guidelines for the AFM algorithm. In the following, a theorem of wide-sense convergence in probability by Benveniste *et al.* [71] is presented. Then the original problem is translated into the study of an associated ODE. Finally, the study of the ODE gives us the desired convergence.

2.3.1 A Theorem of Wide-sense Convergence in Probability

Consider the recursive parameter estimation scheme

$$\hat{\Theta}_{n+1}=\hat{\Theta}_n+\tau V_n(\hat{\Theta}_n), \qquad \hat{\Theta}_0=\hat{\theta}_0\epsilon\hat{D}_c \tag{26}$$

where $\hat{\Theta}_n$ is the parameter vector estimate of dimension d at the n th step. Here, the capital $\hat{\Theta}$ is used to denote the randomness, whereas the lower case $\hat{\theta}$ denotes a deterministic parameter vector. $V_n(\hat{\theta})$ is a stationary random vector function over $\hat{D}_c$ which is in turn a prescribed compact region in $\mathcal{R}^d$. $\tau>0$ is the constant gain. Consider also an associated ordinary differential equation (ODE) with (26):

$$\frac{d\hat{\theta}(t)}{dt}=\nabla[\hat{\theta}(t)], \qquad \hat{\theta}(0)=\hat{\theta}_0. \tag{27}$$

Here $\nabla[\hat{\theta}(t)]=E\{V_n(\hat{\theta})\}|_{\hat{\theta}=\hat{\theta}(t)}$ depends on t but not on n since $V_n(\hat{\theta})$ is stationary. Assume that (27) has a solution $\hat{\theta}^*(t)$, $t\epsilon\mathcal{R}$, which satisfies

$$\hat{\theta}^*(t)\epsilon\hat{D}_c, \qquad t\epsilon\mathcal{R}. \tag{28}$$

Then the following theorem is due to Benveniste *et al.* [71].

Theorem:

If the following assumptions are met:

i) Smoothness. There exists a positive random variable $C_1(\omega)$ such that $E(C_1)<\infty$ and

$$\| V_n(\hat{\theta}_1,\omega)-V_n(\hat{\theta}_2,\omega) \| \leq C_1(\omega) \| \hat{\theta}_1-\hat{\theta}_2 \|, \qquad \text{for all } \hat{\theta}_1,\hat{\theta}_2 \epsilon \hat{D}_c :$$

ii) Boundedness. There exists $C_2<\infty$ such that

$$\left| \sup_{\hat{\theta}\epsilon\hat{D}_c} E \| V_n(\hat{\theta}) \|^2 \right|^{1/2} < C_2:$$

iii) Mixing Condition. There exists a stationary random sequence $\{\xi_n\}_{n\epsilon\mathbb{Z}}$ defined over the probability space $\{\Omega, \mathcal{F}, P\}$ with values in some measurable space $\{E, \mathcal{E}\}$, which is ϕ-mixing in the following sense:

$$\sum_{k=0}^{\infty} \phi_k^{1/2} < \infty$$

where

$$\phi_k = sup \left\{ |P(A \mid B)-P(A)|: \quad \begin{matrix} A \epsilon \sigma\{\xi_l, l \leq 0\} \\ B \epsilon \sigma\{\xi_{n+k}, n \geq 0\} \end{matrix} \right\}.$$

Furthermore, there exists, for every $0 \leq m \leq \infty$, a function $V_n^m(\hat{\theta})$ which is $\mathcal{B}(\hat{D}_c) \otimes \sigma\{\xi_{n-m}, \cdots, \xi_n, \cdots, \xi_{n+m}\}$-measurable, and a decreasing sequence $\{\nu_m\}_{m\geq 0}$ such that

$$\sum_{m=0}^{\infty} \nu_m^{1/2} < \infty$$

where

$$\nu_m \geq \sup_{\theta\epsilon\hat{D}_c} E \| V_n(\hat{\theta})-V_n^m(\hat{\theta}) \|^2.$$

Then, for $S<\infty$ fixed, there exist two positive and finite constants C and C', and a positive function $\epsilon(\tau)$ going to zero with τ, such that the estimates in (26) satisfy

$$P\left\{ \sup_{0\leq n\tau\leq S} \| \hat{\Theta}_n - \hat{\theta}^{\cdot}(n\tau) \| > C\epsilon(\tau) \right\} < C'\epsilon(\tau) \qquad (29)$$

where $\hat{\theta}^{\cdot}(t)$ is given in (28). Furthermore, if $\theta^{\cdot}=\lim_{t\to\infty}\hat{\theta}^{\cdot}(t)$ exists, then

$$P\left\{\| \hat{\Theta}_{[\frac{S}{\tau}-1]}-\theta^{\cdot} \| \geq \eta+C\,\epsilon(\tau)\right\}<C'\epsilon(\tau) \tag{30}$$

where $[\cdot]$ denotes the upper nearest integer and η is such that $\| \hat{\theta}^{\cdot}(S)-\theta^{\cdot} \| <\eta$.

Remark: For adaptive IIR filtering application, the stability is of most importance. Hence, the theorem stated here is actually a modified version of that of [71] in that we are restricted to work in $\hat{D}_c$ instead of $\mathbb{R}^d$. Its importance will be clear in the next section.

The expression of $\epsilon(\tau)$ and other loosely related information will not be covered here. One can refer to [71] for further details.

2.3.2 Association of the Algorithm with the ODE

Using the unit delay operator q^{-1}, i.e., $q^{-1}x(n)=x(n-1)$, the following polynomial operators are introduced:

$$\hat{A}(n,\,q^{-1})=1-\hat{a}_1(n)q^{-1}-\cdots-\hat{a}_{\hat{n}_a}(n)q^{-\hat{n}_a}$$

$$\hat{B}(n,\,q^{-1})=\hat{b}_0(n)+\hat{b}_1(n)q^{-1}+\cdots+\hat{b}_{\hat{n}_b}(n)q^{-\hat{n}_b}$$

$$A(q^{-1})=1-a_1q^{-1}-\cdots-a_{n_a}q^{-n_a}$$

$$B(q^{-1})=b_0+b_1q^{-1}+\cdots+b_{n_b}q^{-n_b}.$$

If the IIR filter and the dynamic plant in Figure 1 are considered as operators, then $\hat{A}(n,\,q^{-1})$ and $\hat{B}(n,\,q^{-1})$ are the denominator and the numerator of the IIR filter, and $A(q^{-1})$ and $B(q^{-1})$ are that of the dynamic plant. Note that their orders $\hat{n}_a$, $\hat{n}_b$, n_a, and n_b are arbitrary.

Let $x(n)$ be the input signal, $v(n)$ be the disturbance (or measurement noise), and

$$\hat{\Theta}_n = \begin{bmatrix} \hat{a}_1(n) \\ \vdots \\ \hat{a}_{\hat{n}_a}(n) \\ \hat{b}_0(n) \\ \vdots \\ \hat{b}_{\hat{n}_b}(n) \end{bmatrix}.$$

$$\phi(n, \hat{\Theta}_n) = \begin{bmatrix} \frac{1}{A(n, q^{-1})} \frac{B(q^{-1})}{A(q^{-1})} x(n-1) + \frac{1}{A(n, q^{-1})} v(n-1) \\ \vdots \\ \frac{1}{A(n, q^{-1})} \frac{B(q^{-1})}{A(q^{-1})} x(n-\hat{n}_a) + \frac{1}{A(n, q^{-1})} v(n-\hat{n}_a) \\ \frac{1}{A(n, q^{-1})} x(n) \\ \vdots \\ \frac{1}{A(n, q^{-1})} x(n-\hat{n}_b) \end{bmatrix}.$$

Then the IF algorithm is (25) is a close approximation (see [70], [75]) of

$$\hat{\Theta}_{n+1} = \hat{\Theta}_n + \tau \phi(n, \hat{\Theta}_n) e(n, \hat{\Theta}_n) \tag{31}$$

where

$$e(n, \hat{\Theta}_n) = \frac{B(q^{-1})}{A(q^{-1})} x(n) + v(n) - \frac{1}{A(n, q^{-1})} \hat{B}(n, q^{-1}) x(n). \tag{32}$$

It is obvious that (31) is in the form of (26) with $V_n(\hat{\Theta}_n) = \phi(n, \hat{\Theta}_n) e(n, \hat{\Theta}_n)$.

Now define the compact regions D_c and $\hat{D}_c$ as closed subsets of stability regions.

$$D_c \subset D_s = \{\theta: \textit{All zeros of } A(z^{-1}) \textit{ are inside of the unit circle}\}$$

$$\hat{D}_c \subset \hat{D}_s = \{\hat{\theta}: \textit{ All zeros of } \hat{A}(z^{-1}) \textit{ are inside of the unit circle } \}. \tag{33}$$

Note that the dimensions of D_c and $\hat{D}_c$ may be different. Usually D_c and $\hat{D}_c$ should be taken as close to D_s and $\hat{D}_s$ as possible. We now have

Theorem 1:

Let $x(n)$, $v(n)$ be stationary processes with finite 1st, 2nd and 4th moments. Assume that

a) $A(q^{-1})$ is stable, i.e., $\theta \epsilon D_c$;

b) $\hat{A}(n, q^{-1})$ is stable for all n, i.e., $\hat{\Theta}_n \epsilon \hat{D}_c$; and

c) $\{x(n), v(n)\}$ is ϕ-mixing in the following sense:

$$\sum_{k=0}^{\infty} \phi_k^{1/2} < \infty$$

where

$$\phi_k = sup \left\{ |P(U|V) - P(U)| : \begin{array}{l} U \epsilon \sigma\{[x(n), v(n)]: n \leq 0\} \\ V \epsilon \sigma\{[x(n+k), v(n+k)]: n \geq 0\} \end{array} \right\}.$$

Then the three conditions in Benveniste's theorem are satisfied, and hence the proposed adaptive IIR filtering algorithm converges to the solution of the ODE (27) and (28) in probability as given in (29) and (30).

To prove Theorem 1, we first introduce two lemmas which are of interest in their own rights. The proofs are contained in the Appendix.

Lemma 1:

Consider the following stochastic linear time-invariant system:

$$Y(n, \hat{\theta}) = F(\hat{\theta}) Y(n-1, \hat{\theta}) + G(\hat{\theta}) U(n) \tag{34}$$

where $F(\hat{\theta})$, $G(\hat{\theta})$ are deterministic bounded matrices, $U(n)$ and $Y(n,\hat{\theta})$ are stationary stochastic processes with $E\{\|U(n)\|\}$ bounded and the initial value $Y(0,\hat{\theta})=Y(0)$ bounded. Furthermore, assume that

i) $F(\hat{\theta})$ has all its eigenvalues inside of the unit circle (which may be translated into constraints on $\hat{\theta}$, e.g., $\hat{\theta}\epsilon\hat{D}_c$); and

ii) $F(\hat{\theta})$ and $G(\hat{\theta})$ are Lipschitz, i.e., there exist $0<C_1, C_2<\infty$ such that[3]

$$\begin{aligned}\|F(\hat{\theta}_1)-F(\hat{\theta}_2)\| &\leq C_1\|\hat{\theta}_1-\hat{\theta}_2\| \\ \|G(\hat{\theta}_1)-G(\hat{\theta}_2)\| &\leq C_2\|\hat{\theta}_1-\hat{\theta}_2\|\end{aligned} \qquad \hat{\theta}_1,\hat{\theta}_2\epsilon\hat{D}_c .$$

Then, $Y(n,\hat{\theta})$ is also Lipschitz in $\hat{\theta}$ for all n, i.e., there exists a random variable $\Upsilon>0$ with $E\{\Upsilon\}<\infty$ such that

$$\|Y(n,\hat{\theta}_1)-Y(n,\hat{\theta}_2)\| \leq \Upsilon\|\hat{\theta}_1-\hat{\theta}_2\| . \tag{35}$$

Lemma 2:

Consider four linear time-invariant IIR filters having stable rational transfer functions. Let $u_i(n)$ denote the ith stationary input and $y_i(n)$ the ith output. Furthermore, assume that $u_i(n)$ has finite 1st, 2nd and 4th moments, i.e.,

$$\begin{aligned}&0\leq|E\{u_i(n)\}|\leq M_1<\infty \\ &0\leq E\{u_i(n)^2\}\leq M_2<\infty \qquad \textit{for all } \begin{matrix}1\leq i\leq 4\\ -\infty<n<\infty\end{matrix} \\ &0\leq E\{u_i(n)^4\}\leq M_4<\infty\end{aligned} \tag{36}$$

Then $y_i(n)$ has finite 1st, 2nd, 3rd and 4th moments, i.e.,

---

[3] For the definition of matrix norms, see [78]. Here we use $\|A\|=[tr.(A^TA)]^{1/2}=[\sum_{i,j}|a_{ij}|^2]^{1/2}$ which is compatible with l_2 vector norm.

$$
\begin{aligned}
&0 \leqslant |E\{y_i(n)\}| \leqslant K_1 < \infty \\
&0 \leqslant |E\{y_i(n)y_j(m)\}| \leqslant K_2 < \infty \\
&0 \leqslant |E\{y_i(n)y_j(m)y_k(\xi)\}| \leqslant K_3 < \infty \\
&0 \leqslant |E\{y_i(n)y_j(m)y_k(\xi)y_l(\zeta)\}| \leqslant K_4 < \infty
\end{aligned} \tag{37}
$$

for all $1 \leqslant i, j, k, l \leqslant 4$ and $-\infty < n, m, \xi, \zeta < \infty$.

Proof of Theorem 1:

The three conditions in Benveniste's theorem are all in terms of the deterministic time-invariant parameter $\hat{\theta} \epsilon \hat{D}_c$ instead of the stochastic time-varying $\hat{\Theta}_n$. This makes the analysis much easier. As shown before, the proposed algorithm can be written as in (26) with $V_n(\hat{\theta}) = \phi(n, \hat{\theta}) e(n, \hat{\theta})$. Because of time-invariance, $e(n, \hat{\theta})$ as given in (32) can be written as

$$
e(n, \hat{\theta}) = \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})} x(n) - \left| \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})} x(n-i) \quad \frac{1}{A(q^{-1})} x(n-j) \right| \hat{\theta} + v(n)
$$

$$
1 \leqslant i \leqslant \hat{n}_a, \quad 0 \leqslant j \leqslant \hat{n}_b
$$

$$
= \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})} x(n) + \frac{1}{A(q^{-1})} v(n)
$$

$$
- \left| \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})} x(n-i) + \frac{1}{A(q^{-1})} v(n-i) \quad \frac{1}{A(q^{-1})} x(n-j) \right| \hat{\theta} \tag{38}
$$

where $\hat{A}(q^{-1}) = 1 - \hat{a}_1 q^{-1} - \cdots - \hat{a}_{\hat{n}_a} q^{-\hat{n}_a}$ with $\hat{\theta} = [\hat{a}_1 \cdots \hat{a}_{\hat{n}_a} \; \hat{b}_0 \cdots \hat{b}_{\hat{n}_b}]^T$ being constant (compare with $\hat{A}(n, q^{-1})$ and $A(q^{-1})$ defined earlier). Define the $(\hat{n}_a + \hat{n}_b + 1)$-dimensional vector

$$Y(n,\hat{\theta})=\begin{bmatrix} y_1(n-i,\hat{\theta}) \\ y_2(n-j,\hat{\theta}) \end{bmatrix}=\begin{vmatrix} \dfrac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)+\dfrac{1}{A(q^{-1})}v(n-i) \\ \dfrac{1}{A(q^{-1})}x(n-j) \end{vmatrix}=\phi(n,\hat{\theta}).$$

$$1\leqslant i\leqslant \hat{n}_a,\quad 0\leqslant j\leqslant \hat{n}_b. \tag{39}$$

We have

$$V_n(\hat{\theta})=\phi(n,\hat{\theta})e(n,\hat{\theta})=Y(n,\hat{\theta})[y_1(n,\hat{\theta})-Y(n,\hat{\theta})^T\hat{\theta}]. \tag{40}$$

i) Smoothness Condition:

It is seen that $Y(n,\hat{\theta})$ is a filtered (by $\dfrac{1}{A(z^{-1})}$) version of $\dfrac{B(q^{-1})}{A(q^{-1})}x(n-i)+v(n-i)$ and $x(n-j)$, and hence can be written as

$$Y(n,\hat{\theta})=F(\hat{\theta})Y(n-1,\hat{\theta})+GU(n)$$

where, assuming $\hat{n}_a=\hat{n}_b$,

$$F(\hat{\theta})=\begin{bmatrix} \hat{a}_1 & \cdots & \hat{a}_{\hat{n}_a} & & 0 & \\ 1 & \ddots & 0 & & & \\ & \ddots & \vdots & & & \\ 0 & & 1 & 0 & & \\ & & & \hat{a}_1 & \cdots & \hat{a}_{\hat{n}_b}\ 0 \\ & 0 & & 1 & \ddots & 0\ 0 \\ & & & & \ddots & \vdots\ \vdots \\ & & & 0 & & 1\ 0 \end{bmatrix} \qquad G=\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}$$

$$U(n)=\begin{vmatrix} \dfrac{B(q^{-1})}{A(q^{-1})}x(n-1)+v(n-1) \\ x(n) \end{vmatrix}.$$

Note that there is no loss of generality in this analysis by assuming $\hat{n}_a=\hat{n}_b$. For

$\hat{n}_b < \hat{n}_a$. $Y(n, \hat{\theta})$ as defined in (39) with $\hat{n}_b = \hat{n}_a$ is longer than what we need. Obviously it is sufficient to show that $Y(n, \hat{\theta})$ for the enlarged $\hat{n}_b$ is Lipschitz. For $\hat{n}_b > \hat{n}_a$. $Y(n, \hat{\theta})$ should be lengthened. $F(\hat{\theta})$ should also be enlarged by appending to it columns of zeros and rows that are similar to the last row here. However, this only adds more zero eigenvalues to $F(\hat{\theta})$, and changes nothing else.

It is clear from the form of $F(\hat{\theta})$ that it has one zero eigenvalue and double eigenvalues which are the roots of $z^{\hat{n}_a}\hat{A}(z^{-1})$ [56, Chapter 2]. By the assumption b), these roots are all inside of the unit circle. It is also easily seen that $F(\hat{\theta})$ is Lipschitz, i.e., there exists $0 \leq C < \infty$ such that

$$\| F(\hat{\theta}_1) - F(\hat{\theta}_2) \| \leq C \| \hat{\theta}_1 - \hat{\theta}_2 \|, \quad \textit{for all } \hat{\theta}_1, \hat{\theta}_2 \in \hat{D}_c$$

because $\| F(\hat{\theta}_1) - F(\hat{\theta}_2) \|^2 = 2\sum_{i=1}^{\hat{n}_a} (\hat{a}_{i(1)} - \hat{a}_{i(2)})^2 \leq 2 \| \hat{\theta}_1 - \hat{\theta}_2 \|^2$. Then by Lemma 1, $Y(n, \hat{\theta})$ is Lipschitz for all n. Now by (40), $\| V_n(\hat{\theta}_1) - V_n(\hat{\theta}_2) \|$ can be written as

$$\| V_n(\hat{\theta}_1) - V_n(\hat{\theta}_2) \| \leq \| Y(n, \hat{\theta}_1) y_1(n, \hat{\theta}_1) - Y(n, \hat{\theta}_2) y_1(n, \hat{\theta}_2) \|$$

$$+ \| Y(n, \hat{\theta}_1) Y(n, \hat{\theta}_1)^T \hat{\theta}_1 - Y(n, \hat{\theta}_2) Y(n, \hat{\theta}_2)^T \hat{\theta}_2 \|$$

$$= \| [10 \cdots 0] Y(n+1, \hat{\theta}_1) Y(n, \hat{\theta}_1)^T - [10 \cdots 0] Y(n+1, \hat{\theta}_2) Y(n, \hat{\theta}_2)^T \|$$

$$+ \| Y(n, \hat{\theta}_1) Y(n, \hat{\theta}_1)^T \hat{\theta}_1 - Y(n, \hat{\theta}_1) Y(n, \hat{\theta}_1)^T \hat{\theta}_2$$

$$+ Y(n, \hat{\theta}_1) Y(n, \hat{\theta}_1)^T \hat{\theta}_2 - Y(n, \hat{\theta}_2) Y(n, \hat{\theta}_2)^T \hat{\theta}_2 \|$$

$$\leq \| Y(n+1, \hat{\theta}_1) Y(n, \hat{\theta}_1)^T - Y(n+1, \hat{\theta}_2) Y(n, \hat{\theta}_2)^T \| + \| Y(n, \hat{\theta}_1) Y(n, \hat{\theta}_1)^T \| \| \hat{\theta}_1 - \hat{\theta}_2 \|$$

$$+ \| Y(n, \hat{\theta}_1) Y(n, \hat{\theta}_1)^T - Y(n, \hat{\theta}_2) Y(n, \hat{\theta}_2)^T \| \| \hat{\theta}_2 \|$$

$$\leq \| Y(n+1, \hat{\theta}_1) \| \| Y(n, \hat{\theta}_1) - Y(n, \hat{\theta}_2) \| + \| Y(n+1, \hat{\theta}_1) - Y(n+1, \hat{\theta}_2) \| \| Y(n, \hat{\theta}_2) \|$$

$$+\| Y(n,\hat{\theta}_1)Y(n,\hat{\theta}_1)^T \| \| \hat{\theta}_1-\hat{\theta}_2 \| + \| Y(n,\hat{\theta}_1) \| \| Y(n,\hat{\theta}_1)-Y(n,\hat{\theta}_2) \| \| \hat{\theta}_2 \|$$

$$+\| Y(n,\hat{\theta}_1)-Y(n,\hat{\theta}_2) \| \| Y(n,\hat{\theta}_2) \| \| \hat{\theta}_2 \|$$

$$\leqslant C(\hat{D}_c,\omega) \| \hat{\theta}_1-\hat{\theta}_2 \|$$

where $E\{C(\hat{D}_c,\omega)\}<\infty$ for all n and for all $\hat{\theta}_1,\hat{\theta}_2 \epsilon \hat{D}_c$ by virtue of the stability assumptions a), b) and again Lemma 1. Thus condition i) is satisfied.

ii) Boundedness Condition:

$$E \| V_n(\hat{\theta}) \|^2 = E \| \phi(n,\hat{\theta}) e(n,\hat{\theta}) \|^2$$

$$= E \left\| \begin{bmatrix} \dfrac{B(q^{-1})}{A(q^{-1})A(q^{-1})} x(n-i) + \dfrac{1}{A(q^{-1})} v(n-i) \\ \dfrac{1}{A(q^{-1})} x(n-j) \end{bmatrix} \cdot \right.$$

$$\left. \left[\frac{B(q^{-1})}{A(q^{-1})} x(n) - \frac{\hat{B}(q^{-1})}{A(q^{-1})} x(n) + v(n) \right] \right\|^2$$

$$= E \left\{ \left[\frac{B(q^{-1})}{A(q^{-1})} x(n) - \frac{\hat{B}(q^{-1})}{A(q^{-1})} x(n) + v(n) \right]^2 \cdot \right.$$

$$\left. \left\{ \sum_{i=1}^{\hat{n}_a} \left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})} x(n-i) + \frac{1}{A(q^{-1})} v(n-i) \right]^2 + \sum_{j=0}^{\hat{n}_b} \left[\frac{1}{A(q^{-1})} x(n-j) \right]^2 \right\} \right\}.$$

Let $\dfrac{B(q^{-1})}{A(q^{-1})} x(n) = w_1(n)$, $\dfrac{\hat{B}(q^{-1})}{A(q^{-1})} x(n) = w_2(n)$, $\dfrac{B(q^{-1})}{A(q^{-1})A(q^{-1})} x(n) = w_3(n)$, $\dfrac{1}{A(q^{-1})} x(n) = w_4(n)$, and $\dfrac{1}{A(q^{-1})} v(n) = v_1(n)$. We have

$$E \| V_n(\hat{\theta}) \|^2 = E\left\{ [w_1(n) - w_2(n) + v(n)]^2 \left\{ \sum_{i=1}^{\hat{n}_a} [w_3(n-i) + v_1(n-i)]^2 + \sum_{j=0}^{\hat{n}_b} w_4(n-j)^2 \right\} \right\}$$

$$= E\left\{ \sum_{i=1}^{\hat{n}_a} \{ [w_1(n) - w_2(n)]^2 + 2[w_1(n) - w_2(n)]v(n) + v(n)^2 \} \cdot \right.$$

$$[w_3(n-i)^2 + 2w_3(n-i)v_1(n-i) + v_1(n-i)^2]$$

$$\left. + \sum_{j=0}^{\hat{n}_b} \{ [w_1(n) - w_2(n)]^2 + 2[w_1(n) - w_2(n)]v(n) + v(n)^2 \} w_4(n-j)^2 \right\}$$

$$\leqslant \sum_{i=1}^{\hat{n}_a} \left\{ E\{[w_1(n) - w_2(n)]^2 w_3(n-i)^2\} + 2\,|E[w_1(n) - w_2(n)]^2 w_3(n-i)v_1(n-i)| \right.$$

$$+ E\{[w_1(n) - w_2(n)]^2 v_1(n-i)^2\} + 2\,|E[w_1(n) - w_2(n)]w_3(n-i)^2 v(n)|$$

$$\left. + 4\,|E[w_1(n) - w_2(n)]w_3(n-i)v(n)v_1(n-i)| \right.$$

$$+2\,|E[w_1(n)-w_2(n)]v(n)v_1(n-i)^2|$$

$$+E\{v(n)^2w_3(n-i)^2\}+2\,|Ew_3(n-i)v(n)^2v_1(n-i)|+E[v(n)v_1(n-i)]^2\Big\}$$

$$+\sum_{j=0}^{\hat{n}_b}\Big\{E\{[w_1(n)-w_2(n)]^2w_4(n-j)^2\}+2\,|E[w_1(n)-w_2(n)]w_4(n-j)^2v(n)|$$

$$+Ew_4(n-j)^2v(n)^2\Big\}.$$

By the assumptions and Lemma 2, we see immediately that all terms above are finite for all $\hat{\theta}\epsilon\hat{D}_c$. Hence condition ii) is also satisfied.

iii) Mixing Condition:

The sequences considered in Benveniste's theorem are all two-sided, whereas in a physically realizable situation as ours all sequences start from $n=0$, i.e., one-sided, as considered in Lemma 1. Obviously it makes no difference in our proof for i) and ii). This is also the case for part iii) because of the equivalence of ϕ-mixing for one-sided and two-sided sequences [79, p.169]. For convenience, we will consider two-sided sequences in this part of the proof. This also gives a slightly more general taste. All one-sided sequences can be thought of as two-sided with zeros for $n<0$.

The sequence $\{\xi_n\}$ here is $\{x(n),v(n)\}$ which is ϕ-mixing by the assumption. Thus we only need to show the existence of the sequence $\{\nu_m\}$. We first define $V_n^m(\hat{\theta})$.

Let $h_1(n,\hat{\theta})$ be the unit pulse response of $\frac{B(z^{-1})}{A(z^{-1})A(z^{-1})}$, $h_2(n,\hat{\theta})$ be the unit pulse response of $\frac{1}{A(z^{-1})}$. Then

$$y_1(n,\hat{\theta})=\sum_{n'=-\infty}^{\infty} h_1(n',\hat{\theta})x(n-n')+\sum_{n'=-\infty}^{\infty} h_2(n',\hat{\theta})v(n-n')$$

$$y_2(n,\hat{\theta})=\sum_{n'=-\infty}^{\infty} h_2(n',\hat{\theta})x(n-n').$$

Define the truncated versions of y_1 and y_2 as

$$y_1^m(n,\hat{\theta})=\sum_{n'=-m}^{m} h_1(n',\hat{\theta})x(n-n')+\sum_{n'=-m}^{m} h_2(n',\hat{\theta})v(n-n')$$

$$y_2^m(n,\hat{\theta})=\sum_{n'=-m}^{m} h_2(n',\hat{\theta})x(n-n').$$

Analogous to (39), we define

$$Y^m(n,\hat{\theta})=\begin{vmatrix} y_1^m(n-i,\hat{\theta}) \\ y_2^m(n-j,\hat{\theta}) \end{vmatrix}. \qquad \begin{matrix} 1\leq i\leq \hat{n}_a \\ 0\leq j\leq \hat{n}_b \end{matrix}$$

Then it is natural to define

$$V_n^m(\hat{\theta})=Y^m(n,\hat{\theta})[y_1^m(n,\hat{\theta})-Y^m(n,\hat{\theta})^T\hat{\theta}].$$

We are now ready to calculate

$$[E\,\|V_n(\hat{\theta})-V_n^m(\hat{\theta})\|^2]^{1/2}=\Big\{E\,\|Y(n,\hat{\theta})y_1(n,\hat{\theta})-Y^m(n,\hat{\theta})y_1^m(n\ \hat{\theta})$$

$$-[Y(n,\hat{\theta})Y(n,\hat{\theta})^T-Y^m(n,\hat{\theta})Y^m(n,\hat{\theta})^T]\hat{\theta}\,\|^2\Big\}^{1/2}$$

$$\leq \Big[2E \,\| Y(n,\hat{\theta}) y_1(n,\hat{\theta}) - Y^m(n,\hat{\theta}) y_1^m(n,\hat{\theta}) \|^2$$

$$+ 2E \,\| Y(n,\hat{\theta}) Y(n,\hat{\theta})^T - Y^m(n,\hat{\theta}) Y^m(n,\hat{\theta})^T \|^2 \,\| \hat{\theta} \|^2 \Big]^{1/2}$$

$$\leq \sqrt{2} \Big[E \,\| Y(n,\hat{\theta}) y_1(n,\hat{\theta}) - Y^m(n,\hat{\theta}) y_1^m(n,\hat{\theta}) \|^2 \Big]^{1/2}$$

$$+ \sqrt{2} \,\| \hat{\theta} \| \Big[E \,\| Y(n,\hat{\theta}) Y(n,\hat{\theta})^T - Y^m(n,\hat{\theta}) Y^m(n,\hat{\theta})^T \|^2 \Big]^{1/2}$$

where the inequality $(a+b)^2 \leq 2(a^2+b^2)$ is used. It is more convenient at this point to lump detailed calculation into a lemma whose proof is also contained in the Appendix.

Lemma 3:

With the assumptions of Theorem 1, there exist $0<\alpha_1,\alpha_2<\infty$ and $0<p<1$ such that

i) $\Big[E \,\| Y(n,\hat{\theta}) Y(n,\hat{\theta})^T - Y^m(n,\hat{\theta}) Y^m(n,\hat{\theta})^T \|^2 \Big]^{1/2} \leq \alpha_1 p^m$, *and*

ii) $\Big[E \,\| Y(n,\hat{\theta}) y_1(n,\hat{\theta}) - Y^m(n,\hat{\theta}) y_1^m(n,\hat{\theta}) \|^2 \Big]^{1/2} \leq \alpha_2 p^m$

for all $\hat{\theta} \epsilon \hat{D}_c$.

It is then obvious that we can let $\nu_m^{1/2} = \sqrt{2}\ (\alpha_1 \max_{\theta \epsilon D_c} \| \hat{\theta} \| + \alpha_2) p^m$ and

$$\sum_{m=0}^{\infty} \nu_m^{1/2} < \infty$$

since $0<p<1$. This concludes our proof.

Remarks:

1) Notice that in the proof many common restrictions are not imposed, e.g., i) $x(n)$ and $v(n)$ do not have to be independent; ii) $v(n)$ may not be white; iii) $A(q^{-1})$

and $B(q^{-1})$ (also $\hat{A}(q^{-1})$ and $\hat{B}(q^{-1})$) do not have to be coprime; and iv) nothing is said about the orders. Thus, this convergence is quite general, although this generality will be somehow reduced when we study $\hat{\theta}^*(t)$ (see Theorem 2 in the next section).

2) ϕ-mixing implies uncorrelation when separation is large (∞), e.g., a sequence of independent variables is ϕ-mixing, an m-dependent sequence is also ϕ-mixing [79]. ϕ-mixing also implies ergodicity, see, e.g., [80]. Note that if $x(n)$ and $v(n)$ are independent of each other, then $\{x(n),v(n)\}$ being ϕ-mixing implies each of $\{x(n)\}$ and $\{v(n)\}$ being ϕ-mixing.

3) Since the estimate $\hat{\Theta}_n$ is restricted to fall into $\hat{D}_c$ for all n, a monitoring device has to be incorporated into the algorithm ((25) or (31)) to project $\hat{\Theta}_n$ back into $\hat{D}_c$ once it gets beyond. This might be computationally expensive and, in fact, is not well solved yet [67]. On the other hand, as long as (28) is satisfied, τ can always be selected such that $C\epsilon(\tau)$ in (29) and (30) is much less than the least distance from $\hat{\theta}^*(t)$ to the boundary of $\hat{D}_s$. Then the probability of the filter becoming unstable can be made arbitrarily small in theory. This is confirmed by computer simulations in Section 2.4.

4) The requirement (28) is essential. What happens if (28) is not satisfied can be seen as follows. Suppose $\hat{\theta}^*(t)$ exits $\hat{D}_c$ at a boundary point $\bar{\theta}$. Then by (29), $\hat{\Theta}_n$ will attempt to exit $\hat{D}_c$ at somewhere near $\bar{\theta}$. However, the projection device [see 4) above] will pull it back into $\hat{D}_c$. Because of (29), $\hat{\Theta}_n$ will try to exit again. An infinite number of repetitions of this process result in $\hat{\Theta}_n$ converging (in probability) to the boundary point $\bar{\theta}$, instead of tracing $\hat{\theta}^*(t)$ further, as also noted in [57, Section 4.3.3]. Satisfaction of the requirement (28) has not yet been well studied, although such difficulty has not been seen in our computer

simulations. However, it is certainly crucial for convergence and should not be omitted as in [70].

5) As pointed out in [71], S in (29) and (30) can never be infinity. This may not be very restrictive since in many applications the algorithm is to be turned off after a finite number of adaptations.

2.3.3 Stability of the ODE

Having established (29) and (30), the next step is naturally to study the ODE (27) and its solution $\hat{\theta}^{*}(t)$. A desirable property is the parameter convergence, i.e., if $\lim_{t\to\infty}\hat{\theta}^{*}(t)=\theta^{*}$ exists, would it be the same as the true parameter vector $\theta=[a_1 \cdots a_{n_a}\ b_0 \cdots b_{n_b}]^T \in D_c$? We consider "sufficient order" case and "reduced order" case separately.

A) Sufficient Order Case:

Sufficient order means $n^{*}=min\{\hat{n}_a-n_a, \hat{n}_b-n_b\}\geq 0$. In this case $D_c \subset \hat{D}_c$. It can then be written $\theta=[a_1 \cdots a_{n_a} 0 \cdots 0\ b_0 b_1 \cdots b_{n_b} 0 \cdots 0]^T \in \hat{D}_c$. For $n^{*}=0$, the global stability of the ODE is guaranteed by the following theorem.

Theorem 2:

Let $n^{*}=0$. Assume that the stationary processes $x(n)$ and $v(n)$ are independent of each other, and $A(z^{-1})$ and $B(z^{-1})$ are coprime. Let the stability conditions a) and b) in Theorem 1 hold (substitute $\hat{\theta}(t)$ for $\hat{\Theta}_n$ in b)).

i) If $v(n)$ is white and $x(n)$ is persistently exciting of order $m+1$ with $m=\hat{n}_a+\hat{n}_b$, i.e., there exist $0<\rho_1,\rho_2<\infty$ such that

$$0<\rho_1 I \leqslant E\begin{bmatrix} x(n) \\ \vdots \\ x(n-m) \end{bmatrix}[x(n)\cdots x(n-m)]\leqslant \rho_2 I . \tag{41}$$

then the ODE (27) is stable and $\hat{\theta}^*(t)$ converges to the true parameter value θ.

ii) If $v(n)$ is not white, then the ODE is not guaranteed to be stable and the estimates (if they converge) are generally biased.

Proof:

For simplicity, we drop the time-variable t in the notation $\hat{\theta}(t)$.

i) Observe the identity

$$\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n)=\left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\quad \frac{1}{A(q^{-1})}x(n-j)\right]\theta,\ 1\leqslant i\leqslant \hat{n}_a, 0\leqslant j\leqslant \hat{n}_b .$$

and consider the RHS of the ODE. Since $x(n)$ and $v(n)$ are mutually independent, we have (using (38))

$$E[V_n(\hat{\theta})]=E[\phi(n,\hat{\theta})e(n,\hat{\theta})]$$

$$=E\begin{bmatrix} \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i) \\ \frac{1}{A(q^{-1})}x(n-j) \end{bmatrix}\left\{\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n)\right.$$

$$\left.-\left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\quad \frac{1}{A(q^{-1})}x(n-j)\right]\hat{\theta}\right\}+E\begin{bmatrix} \frac{1}{A(q^{-1})}v(n-i) \\ 0 \end{bmatrix}v(n)$$

$$=E\left[\begin{array}{c}\dfrac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\\ \dfrac{1}{A(q^{-1})}x(n-j)\end{array}\right]\left[\dfrac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\quad\dfrac{1}{A(q^{-1})}x(n-j)\right](\theta-\hat{\theta})$$

$$+E\left[\begin{array}{c}\dfrac{1}{A(q^{-1})}v(n-i)\\ 0\end{array}\right]v(n) \tag{42}$$

$$=-E\left[\begin{array}{c}\dfrac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\\ \dfrac{1}{A(q^{-1})}x(n-j)\end{array}\right]\left[\dfrac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\quad\dfrac{1}{A(q^{-1})}x(n-j)\right](\hat{\theta}-\theta). \tag{43}$$

The last step is obtained because $v(n)$ is white and $\dfrac{1}{A(q^{-1})}v(n-i)$ depends only on past values of $v(n-i)$. Consider (see [69])

$$E\left[\begin{array}{c}\dfrac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\\ \dfrac{1}{A(q^{-1})}x(n-j)\end{array}\right]\left[\dfrac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\quad\dfrac{1}{A(q^{-1})}x(n-j)\right]$$

$$=\mathcal{S}(A,B)E\left[\begin{array}{c}\dfrac{1}{A(q^{-1})A(q^{-1})}x(n)\\ \vdots\\ \dfrac{1}{A(q^{-1})A(q^{-1})}x(n-m)\end{array}\right]\left[\dfrac{1}{A(q^{-1})A(q^{-1})}x(n)\cdots\right.$$

$$\left.\cdots\dfrac{1}{A(q^{-1})A(q^{-1})}x(n-m)\right]\mathcal{S}(A,B)^T \tag{44}$$

where

$$\mathcal{S}(A,B)=\begin{bmatrix} 0\ b_0\ b_1 \cdots b_{n_b} & & \\ & \ddots & \\ & 0\ b_0\ b_1 \cdots b_{n_b} & \\ 1\ -a_1 \cdots -a_{n_a} & & \\ & \ddots & 0 \\ & 1\ -a_1 \cdots -a_{n_a} & \end{bmatrix} \qquad \text{if } \hat{n}_a > n_a,\ \hat{n}_b = n_b,$$

$$\mathcal{S}(A,B)=\begin{bmatrix} 0\ b_0\ b_1 \cdots b_{n_b} & & 0 \\ & \ddots & \\ & 0\ b_0\ b_1 \cdots b_{n_b} & \\ 1\ -a_1 \cdots -a_{n_a} & & \\ & \ddots & \\ & 1\ -a_1 \cdots -a_{n_a} & \end{bmatrix} \qquad \text{if } \hat{n}_a = n_a,\ \hat{n}_b > n_b,$$

$$\mathcal{S}(A,B)=\begin{bmatrix} 0\ b_0\ b_1 \cdots b_{n_b} & & \\ & \ddots & \\ & 0\ b_0\ b_1 \cdots b_{n_b} & \\ 1\ -a_1 \cdots -a_{n_a} & & \\ & \ddots & \\ & 1\ -a_1 \cdots -a_{n_a} & \end{bmatrix} \qquad \text{if } \hat{n}_a = n_a,\ \hat{n}_b = n_b.$$

Since $A(z^{-1})$ and $B(z^{-1})$ are coprime, the Sylvester matrix $\mathcal{S}(A,B)$ for these three cases are all non-singular [56], [81]. The matrix in the middle of (44) is positive definite for all $\hat{\theta}(t)\epsilon\hat{D}_c$ by persistent excitation assumption and [57, Lemma 4.7]. Thus, from the structure of the matrix (44), we see that it is positive definite for all $\hat{\theta}(t)\epsilon\hat{D}_c$.

We now choose $V=\frac{1}{2}\|\hat{\theta}-\theta\|^2 \geq 0$ as the Lyapunov function of the ODE. The derivative of V is

$$\frac{dV}{dt}=(\hat{\theta}-\theta)^T\frac{d\hat{\theta}}{dt}$$

$$=-(\hat{\theta}-\theta)^T E\left[\begin{array}{c}\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\\ \frac{1}{A(q^{-1})}x(n-j)\end{array}\right]\left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\quad\frac{1}{A(q^{-1})}x(n-j)\right](\hat{\theta}-\theta)$$

$$\leqslant 0 \tag{45}$$

with equality iff $\hat{\theta}=\theta$ at which point $\frac{d\hat{\theta}}{dt}=0$. Thus the assertions in i) are proved.

ii) For $v(n)$ non-white, one cannot derive (43) from (42). Thus, it may not be feasible to find a Lyapunov function. If we consider possible convergence points, i.e., stationary points of the ODE, it is seen that $E[V_n(\hat{\theta})]=0$ results in

$$\hat{\theta}=\theta+\left\{E\left[\begin{array}{c}\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\\ \frac{1}{A(q^{-1})}x(n-j)\end{array}\right]\left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\quad\frac{1}{A(q^{-1})}x(n-j)\right]\right\}^{-1}$$

$$E\left[\begin{array}{c}\frac{1}{A(q^{-1})}v(n-i)\\ 0\end{array}\right]v(n) \tag{46}$$

which does not lead to $\hat{\theta}=\theta$ and hence $\hat{\theta}$ (if exists) will be biased and may even go beyond the region $\hat{D}_s$.

Remarks:

1) Theorem 2 together with Theorem 1 give a rigorous convergence proof for the IF algorithm in the sense of (30) with θ^* replaced by θ.

2) The error surface may not be unimodal even for the sufficient order case [48]. Thus, Theorem 2 proves global convergence regardless of local minima for this case (error surface case 2)).

3) For $n^* > 0$, it was shown in [69] that the stationary points of the ODE are given by $\hat{A}(z^{-1}) = A(z^{-1})L(z^{-1})$ and $\hat{B}(z^{-1}) = B(z^{-1})L(z^{-1})$ where $L(z^{-1})$ is a polynomial of degree n^* with zeros inside the unit circle.

B) Reduced Order Case:

In this case $n^* < 0$. If all conditions in Theorem 2 i) are met, the stationary points will satisfy

$$E\{\phi(n,\hat{\theta})e(n,\hat{\theta})\}$$

$$= E\left[\begin{matrix} \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i) \\ \frac{1}{A(q^{-1})}x(n-j) \end{matrix}\right]\left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n)\right.$$

$$\left. - \left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i) \quad \frac{1}{A(q^{-1})}x(n-j)\right]\hat{\theta}\right]$$

$$= E\left[\begin{matrix} \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-1) \\ \vdots \\ \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-\hat{n}_a) \\ \frac{1}{A(q^{-1})}x(n) \\ \vdots \\ \frac{1}{A(q^{-1})}x(n-\hat{n}_b) \end{matrix}\right]\left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-1)\cdots\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-n_a)\right.$$

$$\left.\frac{1}{A(q^{-1})}x(n)\cdots\frac{1}{A(q^{-1})}x(n-n_b)\right|\theta$$

$$-E\left[\begin{matrix}\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\\ \frac{1}{A(q^{-1})}x(n-j)\end{matrix}\right]\left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\quad\frac{1}{A(q^{-1})}x(n-j)\right]\hat{\theta}=0 \qquad (47)$$

which implies

$$\hat{\theta}=\left\{E\left[\begin{matrix}\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\\ \frac{1}{A(q^{-1})}x(n-j)\end{matrix}\right]\left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\quad\frac{1}{A(q^{-1})}x(n-j)\right]\right\}^{-1}\cdot$$

$$E\left[\begin{matrix}\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-1)\\ \vdots\\ \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-\hat{n}_a)\\ \frac{1}{A(q^{-1})}x(n)\\ \vdots\\ \frac{1}{A(q^{-1})}x(n-\hat{n}_b)\end{matrix}\right]\left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-1)\cdots\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-n_a)\right.$$

$$\left.\frac{1}{A(q^{-1})}x(n)\cdots\frac{1}{A(q^{-1})}x(n-n_b)\right|\theta. \qquad (48)$$

Note that the second matrix in (48) is not symmetric any more since $n^* < 0$. Thus, (48) is now a complicated non-linear equation in $\hat{\theta}$, and little can be said about it at present. However, it should be noted that the matrices in (48) still have a special symmetric form due to the use of $U(n)$ (see Proof of Theorem 1, i)) in ϕ vector instead of $\hat{U}(n)=[\frac{\hat{B}(q^{-1})}{A(q^{-1})}x(n-1)+v(n-1)\quad x(n)]^T$ used in most other algorithms such as

Stearns' algorithm and the SHARF. This is a unique feature characterizing the underlying equation error approach [67], which may contribute to the global convergence as will be seen in the next section.

2.4 Computer Simulation

A number of computer simulations were performed for both AFM algorithm and IF algorithm. For most cases, these two algorithms are indeed indistinguishable from casual observance. The only slightly noticeable difference occurs when the adaptive coefficients migrate slightly outside the stable region. In this case, although the adaptive algorithm itself has the ability to pull them back into the stable region if they are not too far away from the boundary (see Figure 7), the adaptation process becomes unstable so that the adaptive coefficients change more drastically. If, however, τ is selected such that the adaptive coefficients strictly remain inside the stable region, the two algorithms were found to always behave exactly the same for all the computer simulations performed. Since it is redundant to present simulations for both algorithms, only those results using the AFM algorithm will be presented.

Let us consider first the simplest situation: error surface case 1). Figure 4 shows one example of Stearns' experiments [47] for this unimodal error surface case. The transfer function of the dynamic plant is $H_p(z^{-1})=\frac{1}{1-1.2z^{-1}+0.6z^{-2}}$ and that of the filter is $H_f(n,z^{-1})=\frac{1}{1-\hat{a}_1(n)z^{-1}-\hat{a}_2(n)z^{-2}}$. The input $x(n)$ is a unit variance white Gaussian pseudonoise sequence. And the disturbance $v(n)$ is set to zero. The error surface is normalized to one and is shown by the contours. Starting from an arbitrary point $(\hat{a}_1(0), \hat{a}_2(0)) = (-0.8, -0.2)$, convergence is obtained for τ=0.001. It was observed that if τ is too large, the adaptive coefficients will go beyond the stable region

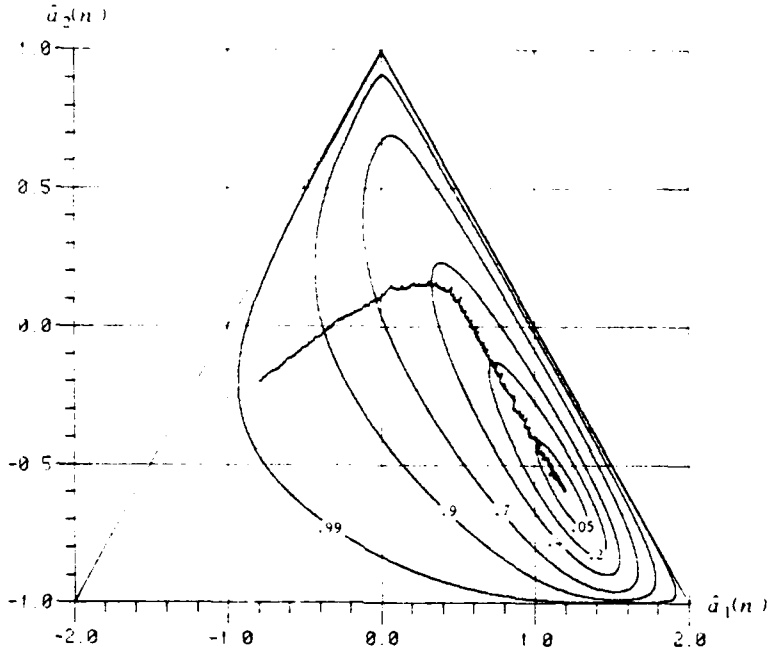


Figure 4 Convergence path for $\hat{a}_1(n)$ and $\hat{a}_2(n)$, sufficient order with white noise input

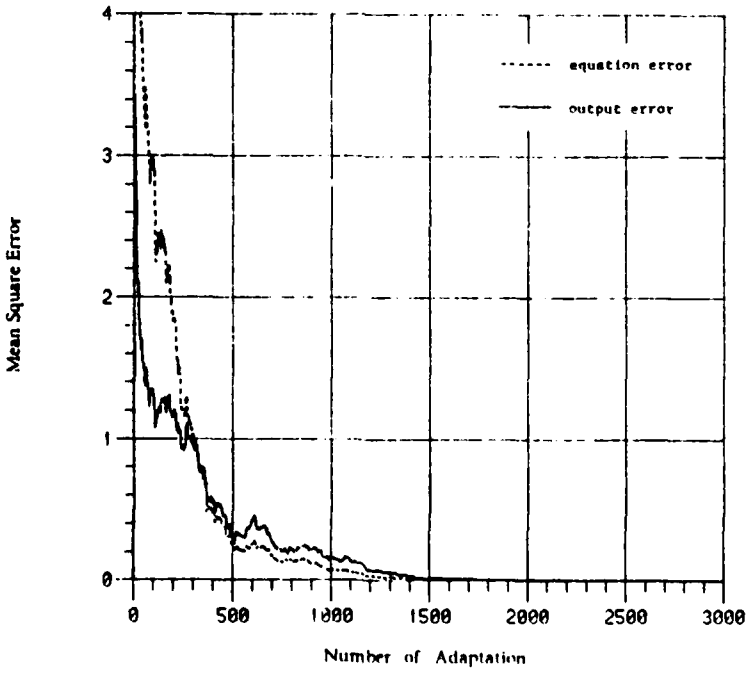


Figure 5 Error progression of Figure 4

and the filter will "blow up." For τ smaller than 0.001 the filter is always stable, and slower convergence results. A plot of the error progression is shown in Figure 5 where, for comparison, the equation error is computed by passing the output error in Figure 3 through the adaptive all-zero post-filter as in Figure 2. It is seen that both the output error and the equation error approach zero after 1500 adaptations.

The same experiment is performed for $v(n)$ a zero mean, unit variance white Gaussian noise sequence. A noisier version of the adaptive coefficient migration is seen in Figure 6. It is seen that at the convergence point, the adaptive coefficients wander around the true value with a larger variance as opposed to Figure 4 due to the non-zero disturbance, and hence demonstrating convergence *in probability* as given in (30). For a plot of error progression, see the dashed lines in Figures 14 - 16.

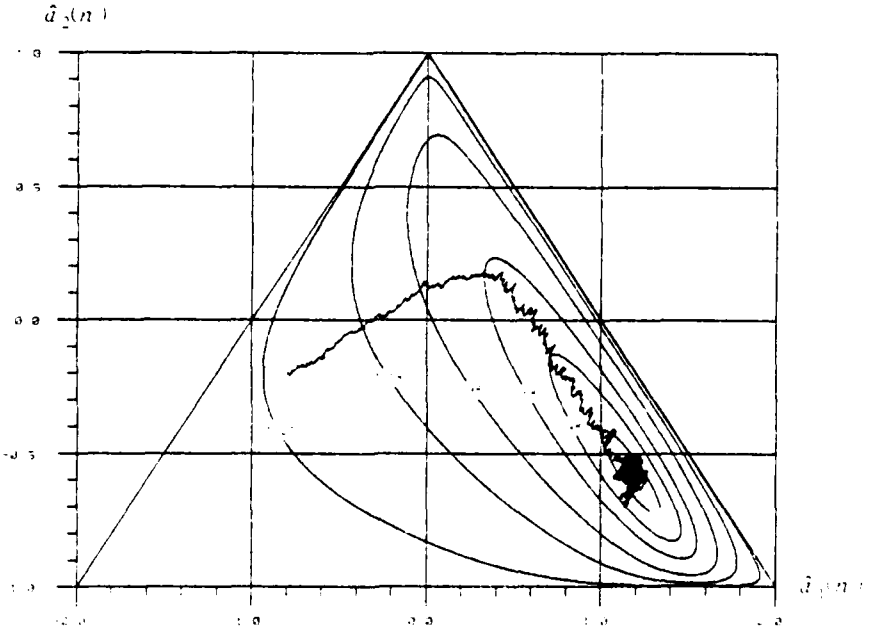


Figure 6 Same as Figure 4 with non-zero disturbance

For the error surface case 2), colored noise input is obtained by filtering a white noise. To ensure the ϕ-mixing requirement and the input richness requirement, the coloring filter used is FIR. The filtered sequence is an m-dependent sequence and its frequency response is never zero. Figure 7 shows a multimodal situation constructed by Soderstrom [48] where $H_p(z^{-1})=\frac{1}{(1-0.7z^{-1})^2}$. $H_f(n,z^{-1})=\frac{\hat{b}_0(n)}{1-\hat{a}_1(n)z^{-1}-\hat{a}_2(n)z^{-2}}$, and the coloring filter is $(1-0.7z^{-1})^2(1+0.7z^{-1})^2$, with $v(n)\equiv 0$. Two minima exist, one being 0.9475 (local) and the other 0 (global). A global convergence path is shown for 4000 iterations starting from near the local minimum point. Note that this is also achievable by SHARF (because the filter orders are sufficient) and LMSEE (because $e(n)=0$ at the global minimum point) but not

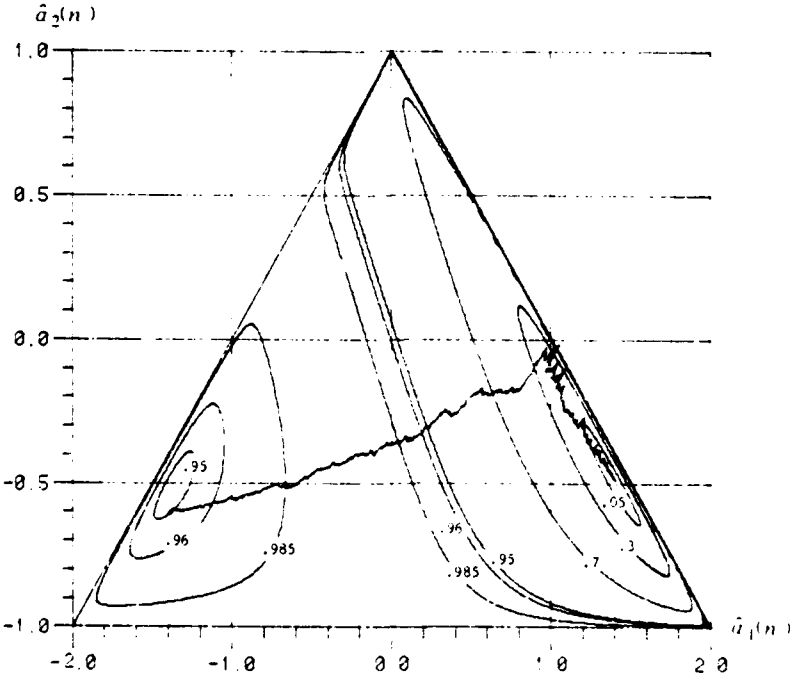


Figure 7 Global convergence path for $\hat{a}_1(n)$ and $\hat{a}_2(n)$, sufficient order with colored noise input

achievable by Stearns' algorithm.

The above examples merely demonstrate the proved convergence in the last section. The results are not surprising also because they can be achieved using some other algorithms as mentioned. However, the examples presented below are unique and interesting, i.e., global convergence regardless of local minima for reduced order cases.

Three examples of case 3) are shown in Figures 8 - 12. Figure 8 shows the example considered by Johnson *et al.* [45] where $H_p(z^{-1})=\dfrac{0.05-0.4z^{-1}}{1-1.1314z^{-1}+0.25z^{-2}}$, $H_f(n,z^{-1})=\dfrac{\hat{b}_0(n)}{1-\hat{a}_1(n)z^{-1}}$, $v(n)\equiv 0$, and the input is again white noise. Again two minima exist (0.976 and 0.277). Starting from *the local minimum*, global convergence is achieved after 6000 iterations (τ=0.0006). This result is remarkable in that it was not

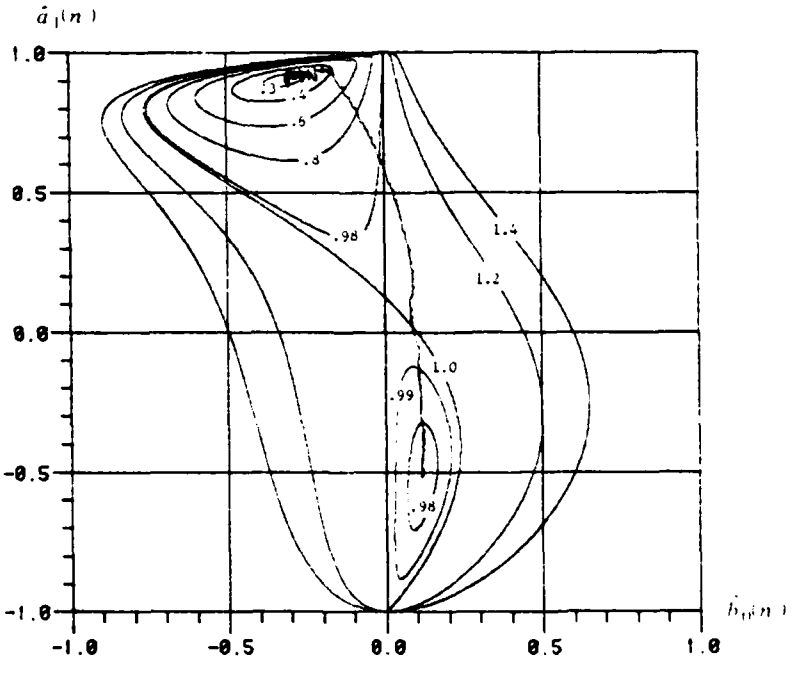


Figure 8 Global convergence path for $\hat{a}_1(n)$ and $\hat{b}_0(n)$, reduced order with white noise input

achieved by any other algorithm [45], [46], [62]. The error progression is shown in Figure 9. It can be seen that minimization of the output error does not lead to minimization of the equation error, and vise versa (compare the two errors at about the 6000th iteration and at about the 3200th iteration). Another example of this case (Stearns, [47]) is shown in Figure 10 where $H_p(z^{-1})=1+10z^{-1}$, $H_f(n,z^{-1})=\dfrac{\hat{b}_0(n)}{1-\hat{a}_1(n)z^{-1}-\hat{a}_2(n)z^{-2}}$, $v(n)\equiv 0$, and the input is white noise. In this case the values at the two minima are very close (0.656 and 0.498). The global convergence is again achieved for more iterations (12000). The same experiment as in Figure 10 is then repeated for $v(n)$ a white Gaussian noise with zero mean and variance 50.5 (normalized to 0.5 [47] in these two figures). Again the same kind of global convergence is observed in Figures 11 and 12.

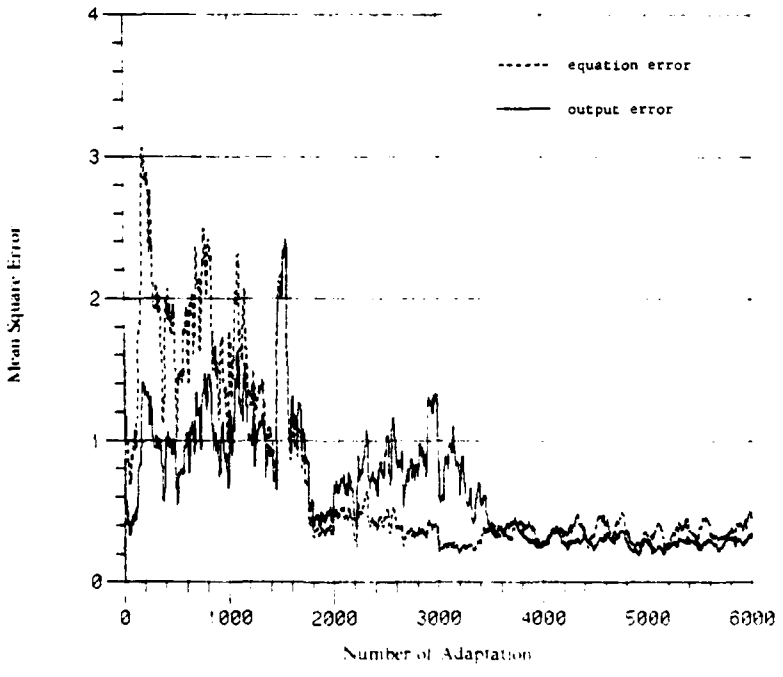


Figure 9 Error progression of Figure 8

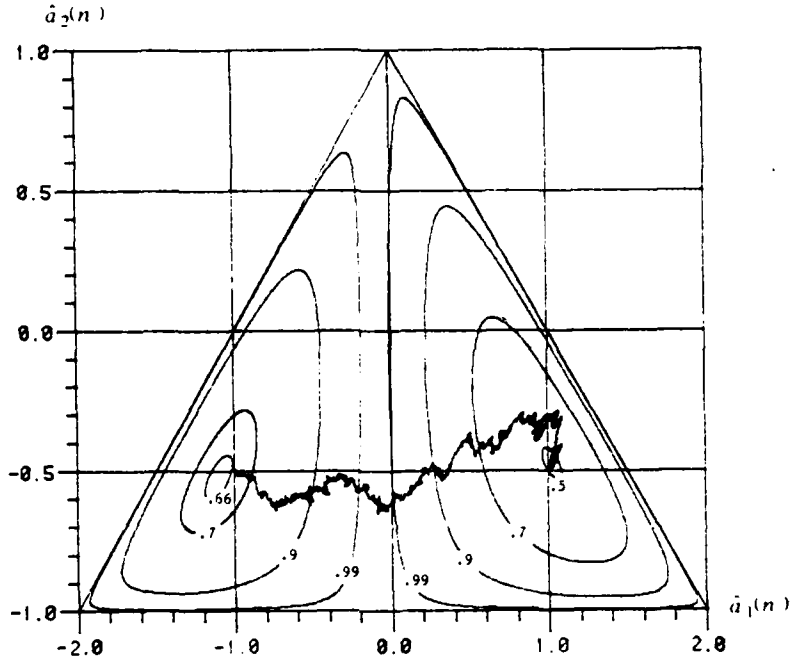


Figure 10 Global convergence path of $\hat{a}_1(n)$ and $\hat{a}_2(n)$, reduced order with white noise input

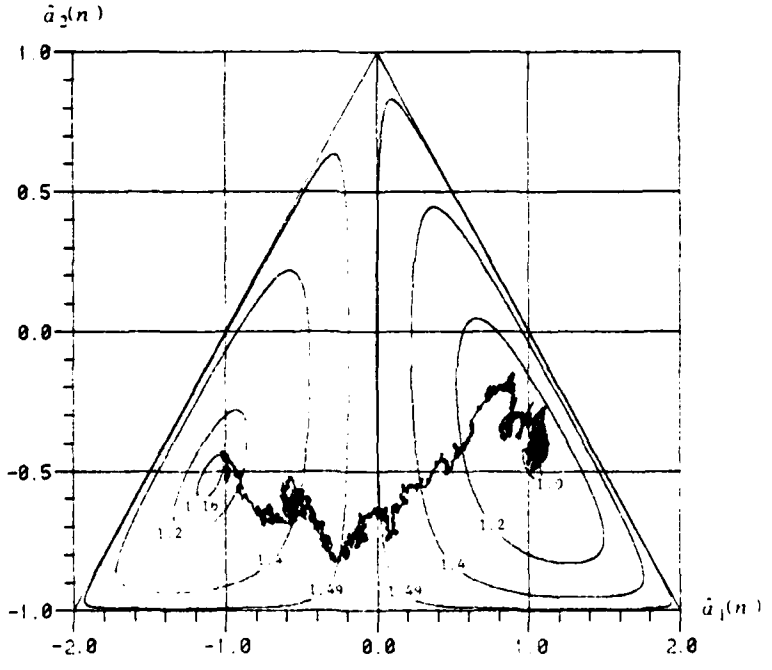


Figure 11 Same as Figure 10 with non-zero disturbance

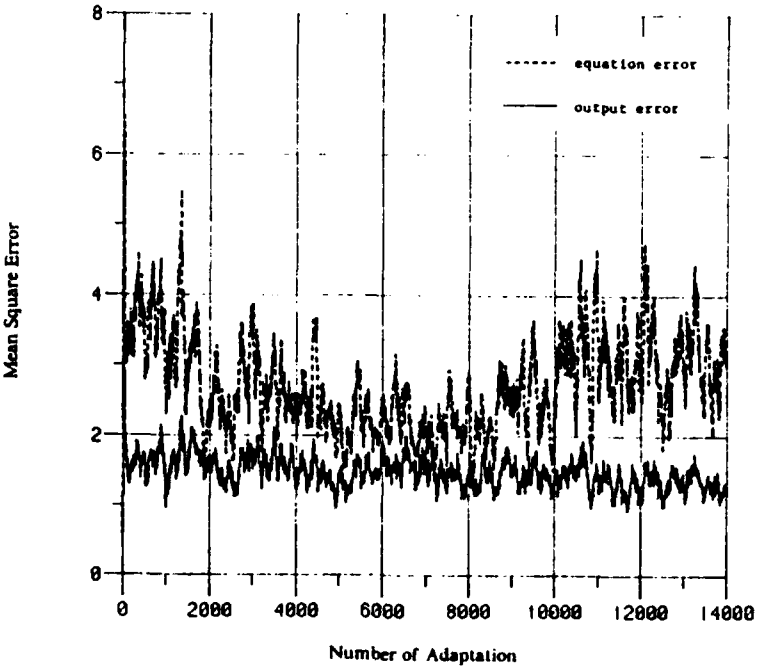


Figure 12 Error progression of Figure 11

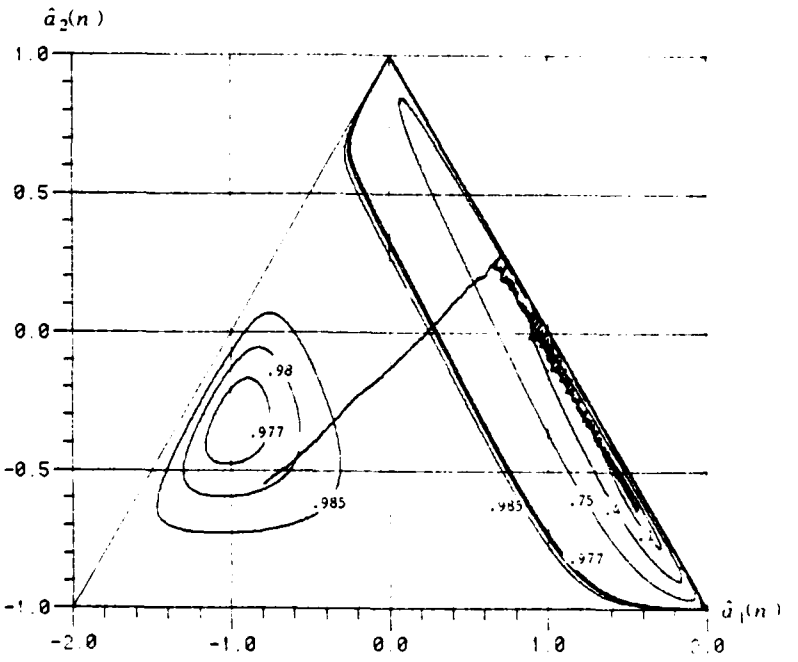


Figure 13 Global convergence path for $\hat{a}_1(n)$ and $\hat{a}_2(n)$, reduced order with colored noise input

Finally, an example of the error surface case 4) is obtained by extending the example in case 2) to $H_p(z^{-1})=\frac{1}{(1-0.6z^{-1})^3}$, $H_f(n,z^{-1})=\frac{\hat{b}_0(n)}{1-\hat{a}_1(n)z^{-1}-\hat{a}_2(n)z^{-2}}$, and the coloring filter $(1-0.6z^{-1})^2(1+0.6z^{-1})^2$, and is shown in Figure 13. Note that the global minimum value is no longer zero (0.016). Global convergence is again observed.

It should be noticed from these simulations that the convergence paths exhibit gradient nature (of the steepest descent method) near convergence points. This confirms the statements below (19).

2.5 Convergence Rate and Other Issues

Two remaining issues concerning the convergence of the IF algorithm are considered in this section: convergence rate and the coloring effects of the disturbance $v(n)$. Again, the fact that the AFM algorithm is a close approximation of the IF algorithm for small τ should also be true for these issues. Here, only sufficient order case will be considered.

2.5.1 Convergence Rate

The convergence rate will be studied for the ideal convergence case only, i.e., when the conditions in Theorem 1 and Theorem 2, i) are met. Assume τ small, then $\hat{\Theta}_n$ closely approximates $\hat{\theta}^*(t)$ in the sense of (29) as proved in Section 2.3. Thus, to study the convergence rate of $\hat{\Theta}_n$, it suffices to investigate the behavior of $\hat{\theta}^*(t)$. From (38) and the identity in the proof of Theorem 2, i) (page 38),

$$e(n,\hat{\theta})=\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n)-\left[\frac{B(q^{-1})}{A(q^{-1})A(q^{-1})}x(n-i)\quad\frac{1}{A(q^{-1})}x(n-j)\right]\hat{\theta}+v(n)$$

$$1 \leqslant i \leqslant \hat{n}_a, \quad 0 \leqslant j \leqslant \hat{n}_b$$

$$= -\left| \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})} x(n-i) \quad \frac{1}{A(q^{-1})} x(n-j) \right| (\hat{\theta}-\theta) + v(n).$$

Since $\{x(n)\}$ and $\{v(n)\}$ are independent.

$$E\{e(n,\hat{\theta})^2\} = E\{v(n)^2\}$$

$$+(\hat{\theta}-\theta)^T E \left| \begin{array}{c} \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})} x(n-i) \\ \frac{1}{A(q^{-1})} x(n-j) \end{array} \right| \left| \frac{B(q^{-1})}{A(q^{-1})A(q^{-1})} x(n-i) \quad \frac{1}{A(q^{-1})} x(n-j) \right| (\hat{\theta}-\theta)$$

$$= -\frac{dV}{dt} + \sigma_v^2 \tag{49}$$

where $V = \frac{1}{2} \| \hat{\theta}-\theta \|^2$, and $\sigma_v^2 = E\{v(n)^2\}$ is the variance of the noise sequence, see (45). Now if $\hat{\theta}(t) = \hat{\theta}^*(t)$, the quantity $V(t)$ then represents the (mean square) parameter error and $E\{e(n,\hat{\theta})^2\}|_{\hat{\theta}=\hat{\theta}^*(t)}$ the (mean square) output error, and they are related by (49). It can be shown that they are both exponential, which was already observed by the computer simulations in Section 2.4. In fact, we can write

$$-\frac{dV}{dt} = [\hat{\theta}^*(t)-\theta]^T R[\hat{\theta}^*(t)][\hat{\theta}^*(t)-\theta] \tag{50}$$

where $R(\hat{\theta})$ is given in (44). By the assumptions and the argument below (44), $R[\hat{\theta}^*(t)]$ is positive definite, i.e.,

$$K_1 I \leqslant R[\hat{\theta}^*(t)] \leqslant K_2 I,$$

and hence

$$K_1 \| \hat{\theta}^*(t)-\theta \|^2 \leqslant -\frac{dV}{dt} \leqslant K_2 \| \hat{\theta}^*(t)-\theta \|^2.$$

i.e.,

$$-2K_1V(t) \geq \frac{dV(t)}{dt} \geq -2K_2V(t). \tag{51}$$

Solving (51) yields the bounds for $V(t)$:

$$V_u(t) = V(0)e^{-2K_1 t} \qquad (Upper\ bound) \tag{52a}$$

$$V_l(t) = V(0)e^{-2K_2 t} \qquad (Lower\ bound). \tag{52b}$$

Because of (49), the bounds for $E\{e[n, \hat{\theta}^*(t)]^2\}$ are similarly

$$E\{e[n, \hat{\theta}^*(t)]^2\}_u = 2K_2V(0)e^{-2K_1 t} + \sigma_v^2 \qquad (Upper\ bound) \tag{53a}$$

$$E\{e[n, \hat{\theta}^*(t)]^2\}_l = 2K_1V(0)e^{-2K_2 t} + \sigma_v^2 \qquad (Lower\ bound). \tag{53b}$$

The time scaling between the ODE (continuous) and the adaptation process (discrete) is $t = n\tau$.

The above bounds are of little practical value, however, because K_1 and K_2 depend on $x(n)$, $\hat{A}^*(q^{-1})$, $A(q^{-1})$, and $B(q^{-1})$ and are difficult to compute in general. For a given set of data $\{x(n)\}$, $\{y(n)\}$, one can only hope to approximate $R[\hat{\theta}^*(t)]$ with a constant matrix and to solve the original ODE as a first order linear differential equation. This approximation might be very rough because $\hat{\theta}^*(t)$ may change considerably during the adaptation. However, locally it can still be accurate and worth investigating.

The simulation example presented in Figure 6, Section 2.4 is used to demonstrate the convergence rate discussed above (for the values used in the example, see page 46). To obtain $V(t)$, a suitable constant approximation for $R[\hat{\theta}^*(t)]$ is made and the ODE as given by (27) and (43) is solved for $\hat{\theta}^*(t)$ [82]:

$$\hat{\theta}^{\cdot}(t)-\theta=e^{-R[\hat{\theta}^{\cdot}(t)](t-t_0)}[\hat{\theta}^{\cdot}(t_0)-\theta]. \tag{54}$$

Then $V(t)=\frac{1}{2}[\hat{\theta}^{\cdot}(t)-\theta]^T[\hat{\theta}^{\cdot}(t)-\theta]$. The mean square output error is calculated using (49) and (50) and plotted on the time scale $t=n\tau$. Three different approximations for $R[\hat{\theta}^{\cdot}(t)]$ are used and their results are shown in Figures 14 - 16. In Figure 14, the matrix $R[\hat{\theta}^{\cdot}(t)]$ is calculated using $A(q^{-1})=1-1.2q^{-1}+0.6q^{-2}$ in place of $\hat{A}(q^{-1})$ and $A(q^{-1})$ (see (44) for the expression of $R(\hat{\theta})$; $B(q^{-1})=1$ in this example). In other words, $\hat{A}(q^{-1})$ is approximated by its final value. Hence, the above calculation should at least be accurate locally around the convergence point, i.e, for t large. Three curves are shown in Figure 14: a sample curve of the mean square output error of the simulated stochastic adaptation process (using $\hat{\Theta}_n$; dashed line), the ensemble average of 20 such curves (solid line), and the predicted convergence curve for the mean square

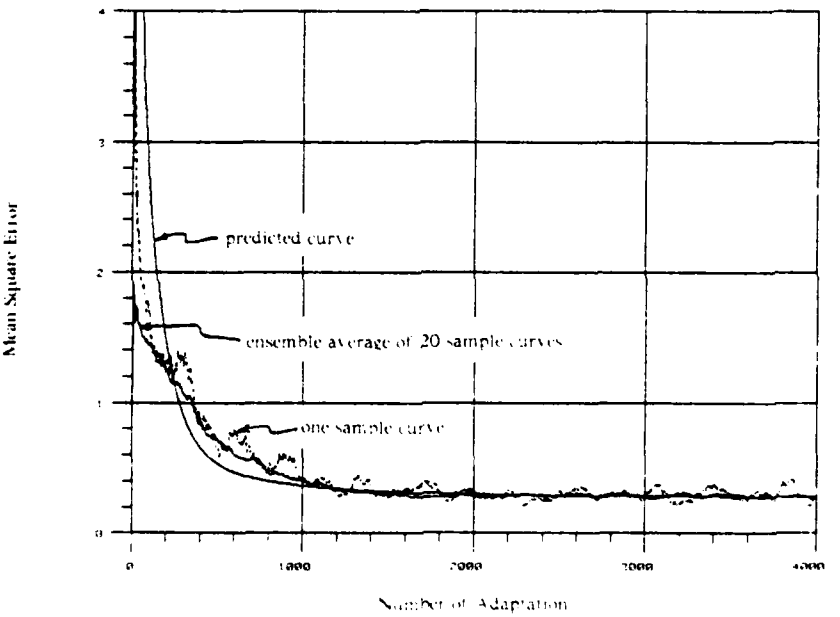


Figure 14 Convergence rate with $\hat{A}(q^{-1})$ approximated by $A(q^{-1})$

error calculated using the above method (solid line). As expected, the predicted curve coincides with the ensemble average fairly well for n large, poorly for n small (at the beginning of adaptation). For the migration of $\hat{\Theta}_n$ see Figure 6. Next, the initial guess made in Figure 6, $\hat{A}(0,q^{-1})=1+0.8q^{-1}+0.2q^{-2}$ is used in place of $\hat{A}(q^{-1})$ while $A(q^{-1})$ is still the same as before, i.e., $\hat{A}(q^{-1})$ is approximated by its initial value. The same group of curves is shown in Figure 15. Again, as expected, the predicted curve coincides with the ensemble average very well at the beginning of the adaptation, and poorly for n large.

Although these two plots are still not practical since $A(q^{-1})$ is unknown in reality, they do give some insight into the rate of convergence for the IF algorithm. A practical approximation for $R[\hat{\theta}^*(t)]$ can be made by replacing the known operator $\hat{A}(0,q^{-1})=1+0.8q^{-1}+0.2q^{-2}$ for both $\hat{A}(q^{-1})$ and $A(q^{-1})$. The result is shown in

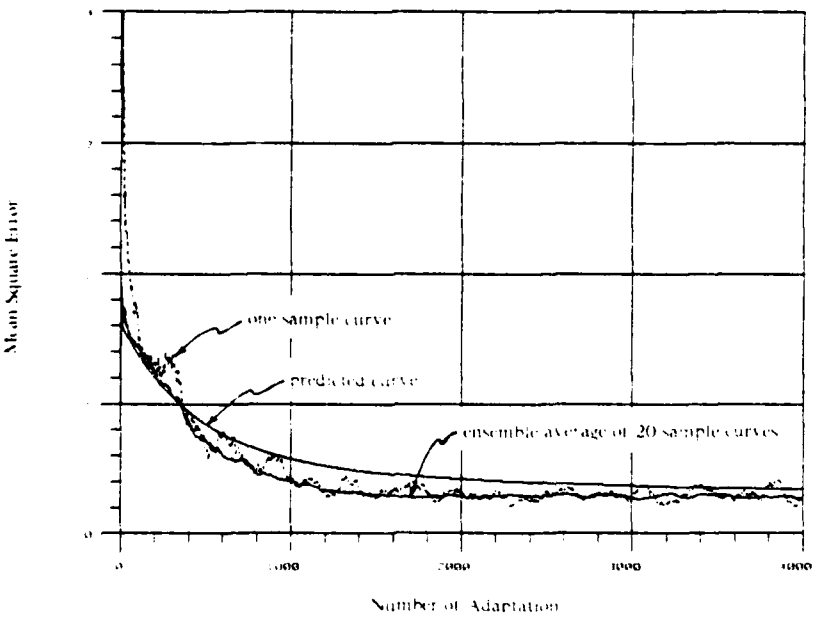


Figure 15 Convergence rate with $\hat{A}(q^{-1})$ approximated by $\hat{A}(0,q^{-1})$

Figure 16, where it is seen that the predicted curve using this approximation is not satisfactory at all. A more accurate approach would be to solve the non-linear ODE without approximating $R[\hat{\theta}^{\cdot}(t)]$, which is inconceivable at this moment.

2.5.2 Coloring Effect

Let the conditions in Theorem 1 and Theorem 2, ii) be met. As pointed out in the proof of Theorem 2, ii), for a non-white disturbance $\{v(n)\}$ it is difficult to analyze the ODE, and the possible convergence points are biased. It is the purpose of this subsection to heuristically discuss some necessary conditions for the bias to remain small.

From (46), it is seen that the parameter bias is caused by an additive vector

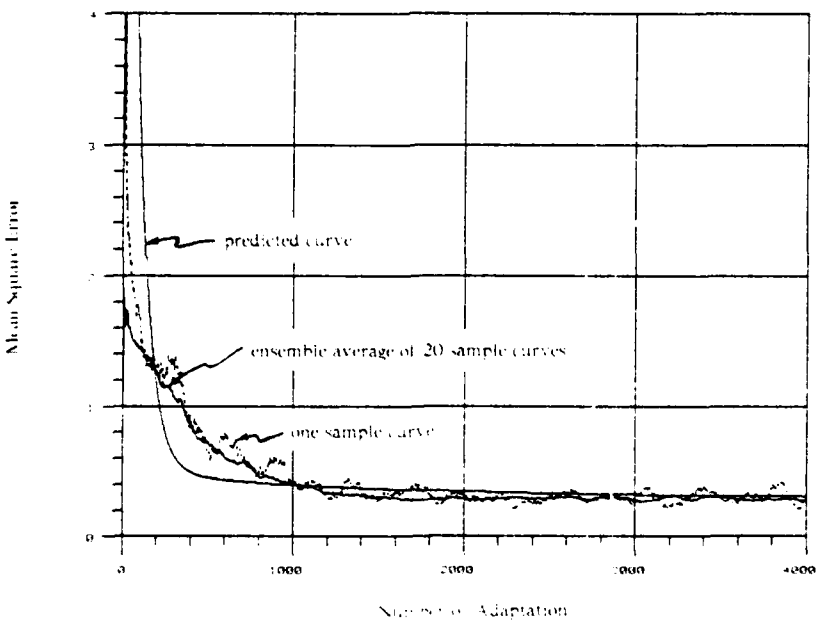


Figure 16 Convergence rate with $\hat{A}(q^{-1})$ and $A(q^{-1})$ approximated by $\hat{A}(0.q^{-1})$

$$\Delta=[R(\hat{\theta})]^{-1}E\left[\begin{matrix}\frac{1}{A(q^{-1})}v(n-i)\\ 0\end{matrix}\right]v(n) \tag{55}$$

where $R(\hat{\theta})$ is given in (44). In order to keep $\|\Delta\|$ small, it is necessary that $\|[R(\hat{\theta})]^{-1}\|$ be small. Using (44), the following can be written:

$$\|[R(\hat{\theta})]^{-1}\| = \|\mathcal{S}(A,B)^{-T}\left\{E\left[\begin{matrix}\frac{1}{A(q^{-1})A(q^{-1})}x(n)\\ \vdots \\ \frac{1}{A(q^{-1})A(q^{-1})}x(n-m)\end{matrix}\right]\left[\frac{1}{A(q^{-1})A(q^{-1})}x(n)\cdots\right.\right.$$

$$\left.\left.\cdots\frac{1}{A(q^{-1})A(q^{-1})}x(n-m)\right]\right\}^{-1}\mathcal{S}(A,B)^{-1}\|$$

$$\leq \|\mathcal{S}(A,B)^{-T}\|\,\|\left\{E\left[\begin{matrix}\frac{1}{A(q^{-1})A(q^{-1})}x(n)\\ \vdots \\ \frac{1}{A(q^{-1})A(q^{-1})}x(n-m)\end{matrix}\right]\left[\frac{1}{A(q^{-1})A(q^{-1})}x(n)\cdots\right.\right.$$

$$\cdots \frac{1}{A(q^{-1})A(q^{-1})} x(n-m) \Bigg] \Bigg|^{-1} \| \| \mathcal{S}(A,B)^{-1} \| . \tag{56}$$

Consider the factor in the middle first. The matrix in the bracket is at least positive semi-definite because of its structure. In order to be invertible, it has to be positive definite and hence the persistent excitation condition (41) must be imposed on the input $\{x(n)\}$ as argued before [57, Lemma 4.7]. This is the same requirement as for the white disturbance case (Theorem 2, i)).

Next, consider the first and the third factor in (56). As mentioned before and proved in [56] and [81], $\mathcal{S}(A,B)$ is non-singular if and only if $A(q^{-1})$ and $B(q^{-1})$ are relatively prime. Moreover, following [56, Theorem A.4.1] it can be shown that for $\| \mathcal{S}(A,B)^{-1} \|$ to be small, the zeros of $A(q^{-1})$ should not even be close to the zeros of $B(q^{-1})$, and vice versa. Let γ be a root of $A(q^{-1})$, and $\gamma+\epsilon$ be a root of $B(q^{-1})$:

$$A(q^{-1})=(1-\gamma q^{-1})(1-a'_1 q^{-1}- \cdots -a'_{n_a-1} q^{-n_a+1})$$

$$B(q^{-1})=(1-\gamma q^{-1}-\epsilon q^{-1})(b'_0+b'_1 q^{-1}+ \cdots +b'_{n_b-1} q^{-n_b+1}).$$

Eliminating $(1-\gamma q^{-1})$ gives

$$A(q^{-1})(b'_0+b'_1 q^{-1}+ \cdots +b'_{n_b-1} q^{-n_b+1})-B(q^{-1})(1-a'_1 q^{-1}- \cdots -a'_{n_a-1} q^{-n_a+1})$$

$$=\epsilon q^{-1}(1-a'_1 q^{-1}- \cdots -a'_{n_a-1} q^{-n_a+1})(b'_0+b'_1 q^{-1}+ \cdots +b'_{n_b-1} q^{-n_b+1}).$$

Equating coefficients of q^{-l} on both sides gives

$$\mathcal{S}(A,B)^T \theta = \epsilon H' \tag{57}$$

where

$$\theta' = [\,-1 \;\; a_1' \cdots a_{n_a-1}' \;\; 0 \;\; b_0' \cdots b_{n_b-1}']^T$$

$$H' = [\,0 \;\; 0 \;\; b'_0 \;\; b'_1 - a'_1 b'_0 \cdots -a'_{n_a-1} b'_{n_b-1}]^T .$$

Taking the norm on both sides of (57) gives

$$\| \theta' \| = \epsilon \| \mathcal{S}(A, B)^{-T} \| \, \| H' \| . \tag{58}$$

Now $\| \theta' \|$ and $\| H' \|$ are of moderate value. Hence, if ϵ is small, i.e., the zeros of $A(q^{-1})$ are close to the zeros of $B(q^{-1})$, $\| \mathcal{S}(A, B)^{-T} \|$ has to be large, thus proving the assertion. Note that this is actually a sensitivity criterion, i.e., the closer the zeros of $A(q^{-1})$ are to the zeros of $B(q^{-1})$, the larger the bias if $v(n)$ is colored.

The two necessary conditions derived, i.e., $\{x(n)\}$ being persistently exciting and the plant not having close pole-zero pairs, are very common in other situations. For example, if the plant has close pole-zero pairs, it would not be surprising to see behavioral difficulties in parameter convergence for any adaptation algorithm.

It should be pointed out that if the correlation of the disturbance is known *a priori*, pre-whitening filters can be used to decorrelate the colored noise as proposed in [83].

3. AN APPLICATION TO ECHO CANCELLATION

The reason for the rapid development of the adaptive filtering, as stated in Chapter 1, is that it has a great potential in wide range of applications due to its unique advantage: adaptability to the (changing) environment. The areas of application for adaptive filters are tremendous [4] - [29]. For adaptive FIR filtering, the usual objective for most applications is the minimization of the mean square output error to obtain the Wiener solution. This proves to be a reasonable objective and is practically useful. For adaptive IIR filtering, however, more is desired. Since an IIR structure can better model a real physical system, one of the purposes for using adaptive IIR filters is to further minimize the mean square output error. Moreover, for the sufficient order case, parameter identification is often desired, in which case the mean square output error can be reduced to zero. However, it can also happen that one can achieve a reasonably small mean square error with a relatively large error in parameter identification, due to the complexity of the error surfaces. One major drawback of the proposed algorithms, as pointed out in the last chapter, is the parameter bias in the presence of a colored disturbance. If parameter identification is desired, this drawback may restrict the algorithms for some applications. However, if only minimization of the mean square output error is necessary, the proposed algorithms still have a large variety of practical applications. On the other hand, there are many situations where the disturbance can be absent, or can be modeled as a white noise sequence. In these cases the proposed algorithms will achieve both minimization of the mean square error and parameter identification in the sufficient order case. For the reduced order case, the proposed algorithms have the potential of achieving global convergence as opposed to other existing algorithms. An important application of this case, echo cancellation in long-distance telephone systems, will be studied in this chapter.

3.1 The Principle of Echo Cancellation

The problem of echo in long-distance telephone systems becoming of great concern to engineers can be dated to more than a half-century ago [84]. Some classic methods were developed which effectively eliminated this problem for systems having a time delay of less than 100 ms at that time. Recently, due to the development of satellite communication technique, larger time delay and hence more severe echoes provoked new blooming for research activities in echo cancellation [13] - [19], [41] - [42], [84] - [85]. The employment of echo cancelers has proved necessary in expanding telephone networks to provide a truly world-wide service.

The cause of echo in a typical 4-wire telephone system [85] is demonstrated in Figure 17. It is a symmetric system for the telephone user 1 and the user 2. Suppose

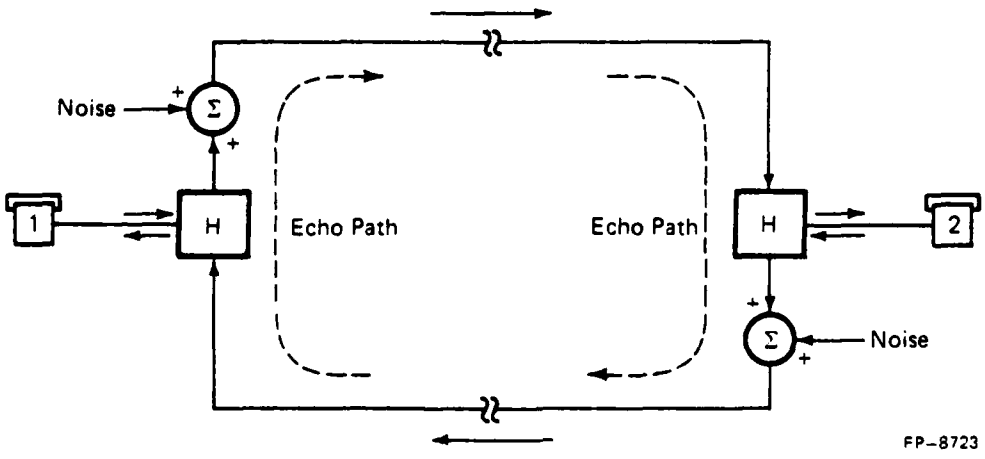


Figure 17 A 4-wire long-distance telephone system

user 1 is speaking and user 2 is listening. The speech signal of user 1 is first transmitted from his telephone set through a 2-wire line to a local office. In the local office a hybrid circuit H which acts as a bridge between 2-wire (bidirectional) and 4-wire (unidirectional) lines connects this speech signal to the upper path of the 4-wire line. The signal is then transmitted over a long distance by the 4-wire line through the upper path. The additive noise represents the noise incurred in the 2-wire line and at the hybrid circuit, which can be well modeled as a white noise. At the receiving end, the hybrid H at a local office near user 2 sends the speech signal to the 2-wire line which connects the telephone set of user 2. Ideally, all the energy of the incoming speech signal from the upper path of the 4-wire line should be transmitted to the 2-wire line and to the user 2. This is done to a large extent in practice by careful design of the sophisticated hybrid circuit H. However, due to the variation of the number of customers and of the length of the 2-wire lines, H can never be designed to match this perfectly. In other words, the transfer function of the path across H along the 4-wire line denoted in Figure 17 as "echo path," is not zero. Thus, the speech signal of user 1 will leak out at the receiving end and will be transmitted back through the lower path of the 4-wire line, which normally transmits the speech signal of user 2. Since the receiving principle for the H at the left is the same as that at the right, user 1 will hear the delayed version of his own speech. This is called talker echo [41], [84] - [85]. The larger time delay on the 4-wire line (i.e., the farther the transmission distance), the more severe and annoying the echo will be.

The user 2 in the above scenario may also hear echo, which is called listener echo [41], [84] - [85]. This echo occurs when the talker echo signal traveling to the left along the lower path of the 4-wire line leaks out along the echo path on the left of the transmission system due to non-zero transfer function at that side. This echo signal is superimposed (time delayed) onto the main speech transmission stream on the upper

path of the 4-wire line, and will be heard by the listener. The listener echo is usually less severe than the talker echo, due to the transmission loss and the loss along the echo paths (the average loss along the echo path is 6 dB [85]). The listener echo can again leak out along the echo path on the right of the transmission system and be transmitted back to user 1, which becomes 2nd talker echo. This process can go on to result in 2nd listener echo, 3rd talker echo, etc.. However, due to the losses, these high order echoes are not of much concern.

It is observed that the transmitted speech signal travels in the 4-wire line for only one-way, whereas echo signals travel for at least a round-trip. Thus, an early method for echo control is to insert a 3 dB loss in each transmission direction in the 4-wire line [85]. Although this method suppresses echo by at least 6 dB, it also introduces an extra 3 dB loss to the speech transmission. A better method using "echo suppressors" were then developed which disrupts transmission in one direction in the 4-wire line if speech is transmitted in the other direction [42], [85]. The echo suppressors are quite effective for line delays less than 100 ms. For larger delay, e.g., in a satellite transmission system, the clipping and the disruption become annoying. It is then necessary to use adaptive echo cancelers.

An adaptive echo canceler is an adaptive filter placed in parallel to the echo path (Figure 18). Its task is to simulate the transfer function of the echo path. By subtracting the estimated returning echo $\hat{y}(n)$ from the real returning echo $y(n)$, the new returning echo $e(n)$ is suppressed greatly. A figure of merit for the effectiveness of an echo canceler is the echo return loss enhancement (ERLE) which is defined as follows [17] - [19]:

$$ERLE = 10 \log \frac{E\{y(n)^2\}}{E\{e(n)^2\}}. \tag{59}$$

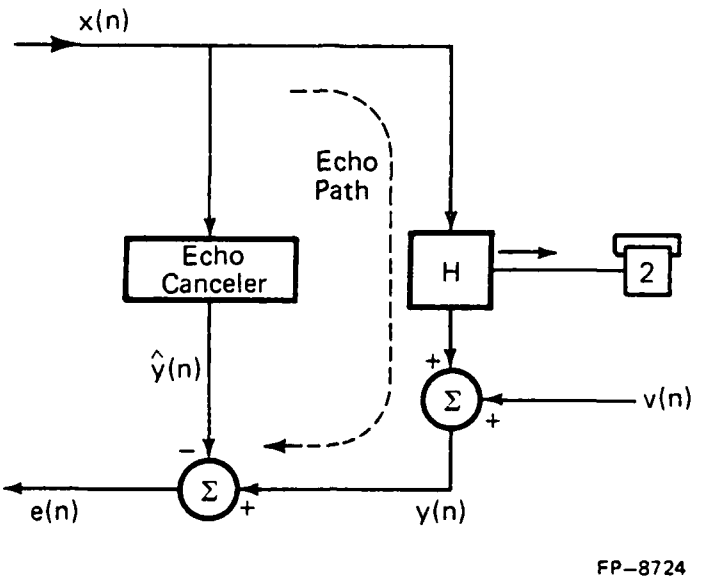


Figure 18 Echo cancellation using an echo canceler

Usually 20 dB or more ERLE is expected for effective echo cancellation.

Almost all echo cancelers today are adaptive FIR filters. Although it is simple to implement, they cannot model the transfer functions of the echo path well. In order to achieve satisfactory performance, the number of weights in the cancelers must be large (e.g., 128 weights for satellite transmission, [85]), which means high cost and high computational complexity. This motivates the investigation of adaptive IIR echo cancelers.

3.2 Adaptive IIR Echo Cancelers

Referring to Figure 18, it is assumed that user 2 is quiet. In case of double-talking (when user 2 tries to interrupt), it is assumed that there is some mechanism (speech

detector) to freeze the adaptation process [85]. If the transfer function of the echo path is considered as an unknown dynamic plant, then the system described in Figure 18 is exactly the same as that in Figure 1. Since the transfer function of the echo path can be well modeled as a rational function [85], it is much more suitable to use an adaptive IIR filter to simulate or to identify it than an adaptive FIR filter. The proposed algorithm and the previous analysis can be applied to the echo cancellation as depicted in Figure 18 immediately. A number of computer simulations for various situations are conducted and will be presented in this section. In order to save computation, the AFM algorithm given by (24) is used throughout this section. As mentioned before, due to slow adaptation the AFM algorithm is a close approximation of the IF algorithm whose convergence behavior is well studied in the last chapter. Rational functions are used to model the echo path transfer function. The AFM algorithm is implemented and compared with SHARF and an adaptive FIR canceler. The disturbance $\{v(n)\}$ is assumed to be a white noise sequence. The results will be presented for sufficient order case and reduced order case respectively.

3.2.1 Sufficient Order Case

An echo path transfer function usually exhibits an all-pass characteristic [85]. Hence, the echo path in this case is modeled as a second order all-pass filter with two complex conjugate poles located at $0.95e^{\pm j30^0}$ and two zeros at $1.0526e^{\pm j30^0}$. The gain is selected as 0.44 such that the loss for all frequencies is about 6 dB, as mentioned before. Figure 19 shows the magnitude of the frequency response and Figure 20 shows the phase. The canceler is a second order adaptive IIR filter with five adaptive coefficients $\hat{a}_1(n)$, $\hat{a}_2(n)$, and $\hat{b}_0(n)$ - $\hat{b}_2(n)$. Three kinds of input signals are used: a digitized speech signal of the sentence "you had a problem with the blush" by a male

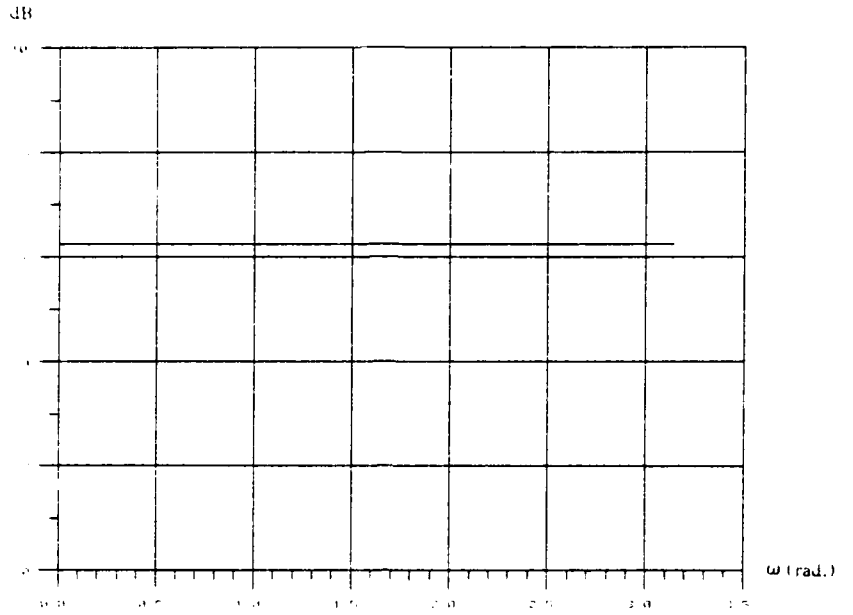


Figure 19 Magnitude (loss) of the 2nd order echo path transfer function

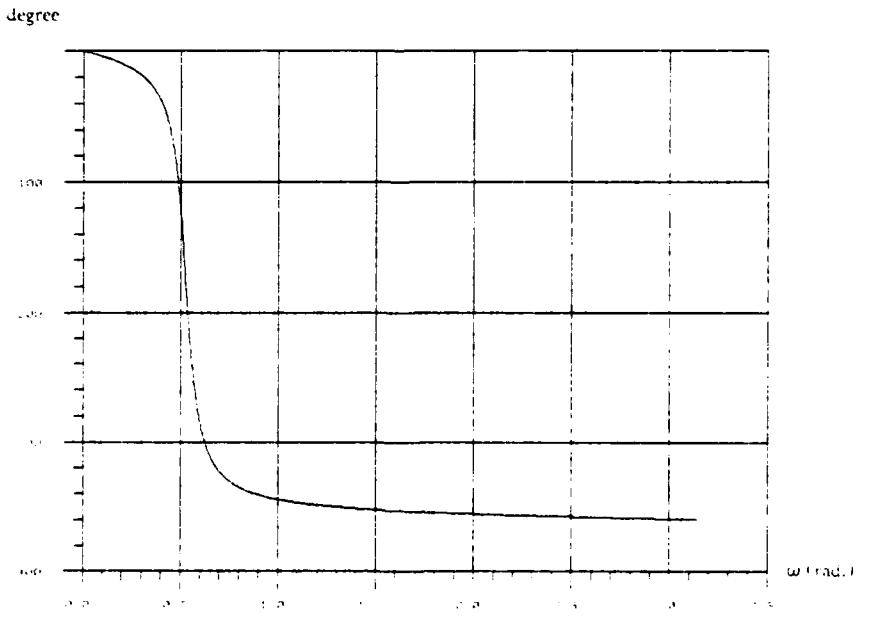


Figure 20 Phase of the 2nd order echo path transfer function

voice (Figure 21), a digitized babble signal uttered by many speakers simultaneously (Figure 22), and a zero mean, unit variance stationary white Gaussian noise sequence. The sampling frequency for the speech and the babble is 20 kHz. Note that the babble signal is essentially a band-limited noise. In order to cope with the drastic magnitude change of the speech signal, the AFM algorithm is implemented with gain normalization, i.e., τ is replaced by $\dfrac{\tau}{\sum_{i=1}^{n_a} y'(n-i)^2+\sum_{j=0}^{n_b} x'(n-j)^2}$ (see (24)). This regulates the convergence speed quite effectively but does not change the convergence properties much, although extra computation is required. Simulation results for $v(n)\equiv 0$ will be presented first, in which case all adaptive coefficients will start from zero initial values.

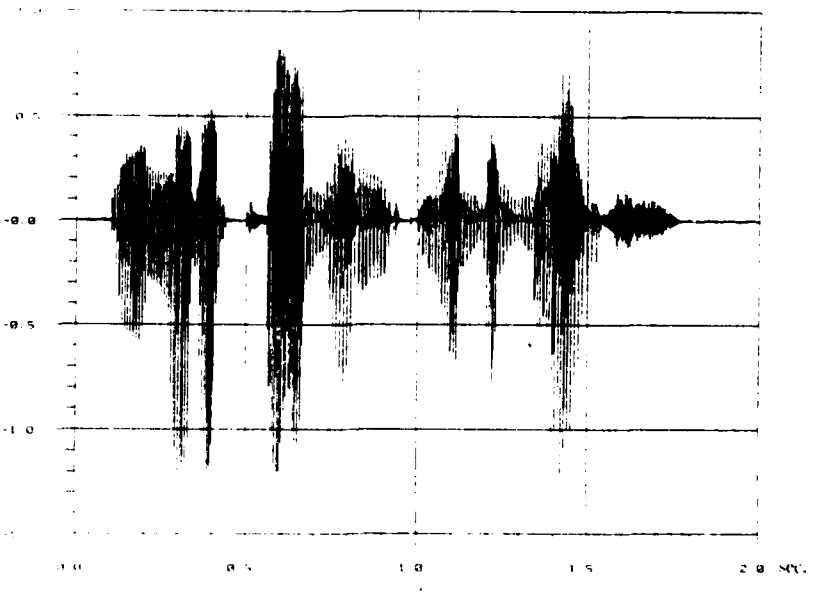


Figure 21 Digitized speech signal

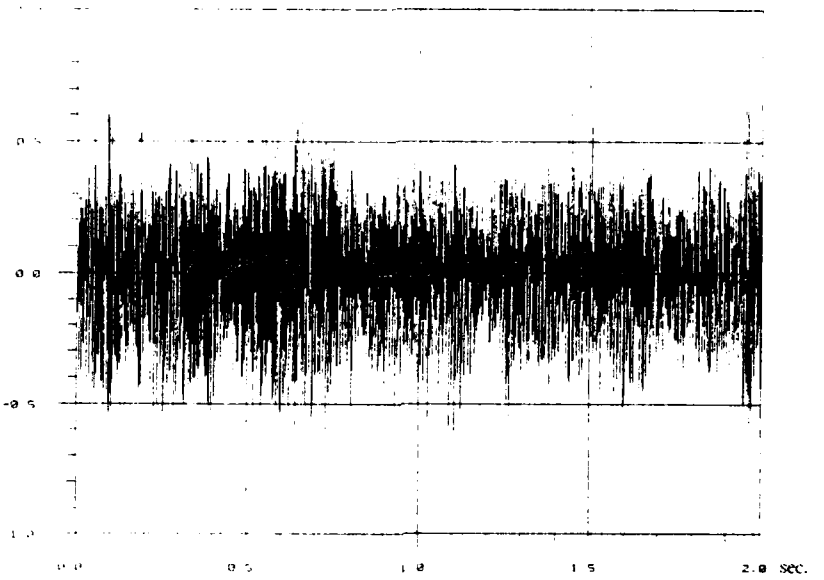


Figure 22 Figure 22 Digitized babble signal

The first group of plots shows the basic simulation results using the adaptive IIR echo canceler for the three different kinds of input. For the speech signal as the input, Figure 23 shows the decrease of the mean square returning echo $E\{e(n)^2\}$ as in Figure 18, and Figure 24 shows the convergence path for the two adaptive coefficients in the denominator. The error surface contour is not analyzed for this case due to the complexity of the speech signal. Only stability boundary is drawn in Figure 24. From the pole location of the plant the values of the coefficients of the denominator are easily calculated as a_1=1.6454 and a_2=−0.9025. Convergence of the adaptive coefficients to this point is achieved after 50,000 adaptations. Figure 25 shows the returning echo for the babble input, and Figure 26 shows the convergence path for the same adaptive coefficients as in Figure 24. Figures 27 and 28 are the same but for the white noise input. The value τ=0.04 is used for all these three cases. It can be noticed that the

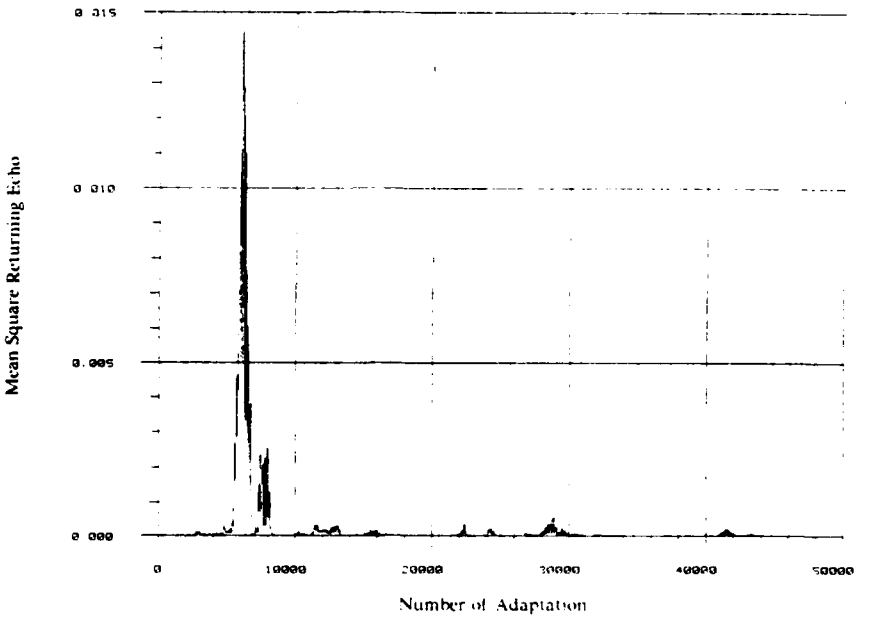


Figure 23 Mean square returning echo for speech input

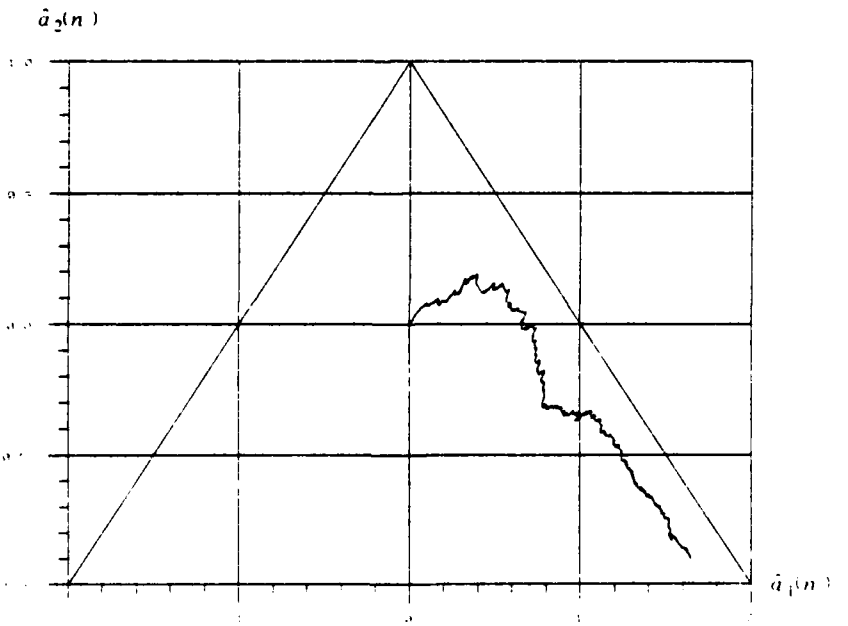


Figure 24 Convergence path of the denominator for speech input

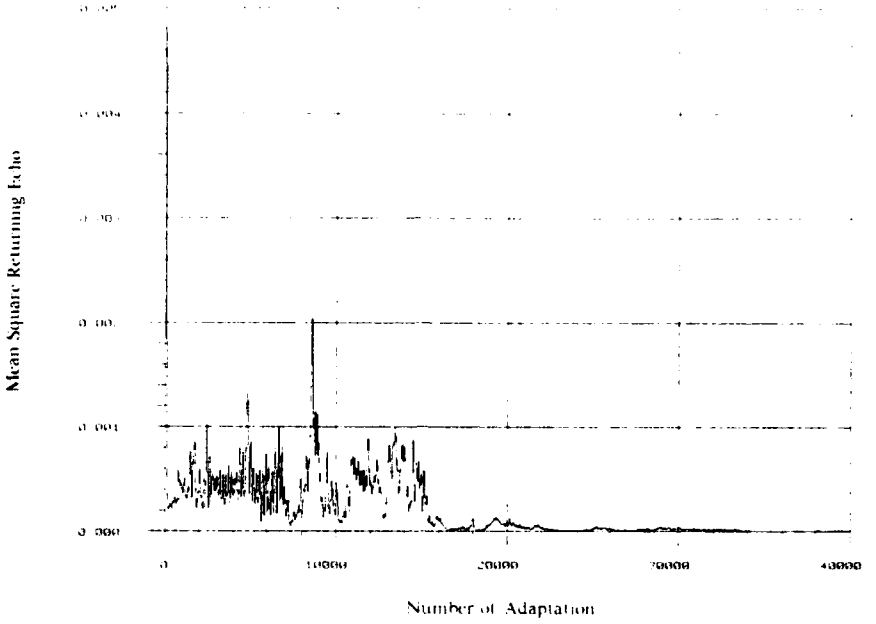


Figure 25 Mean square returning echo for babble input

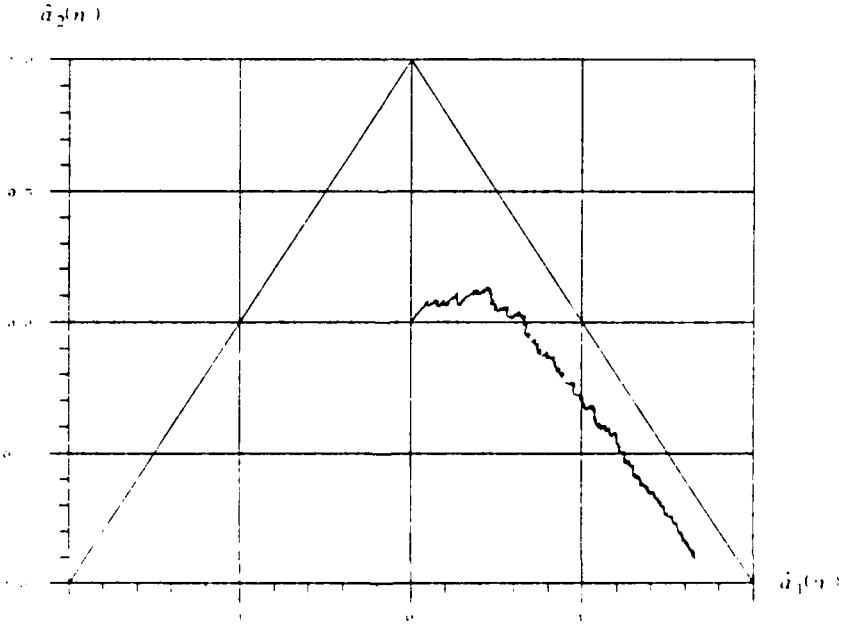


Figure 26 Convergence path of the denominator for babble input

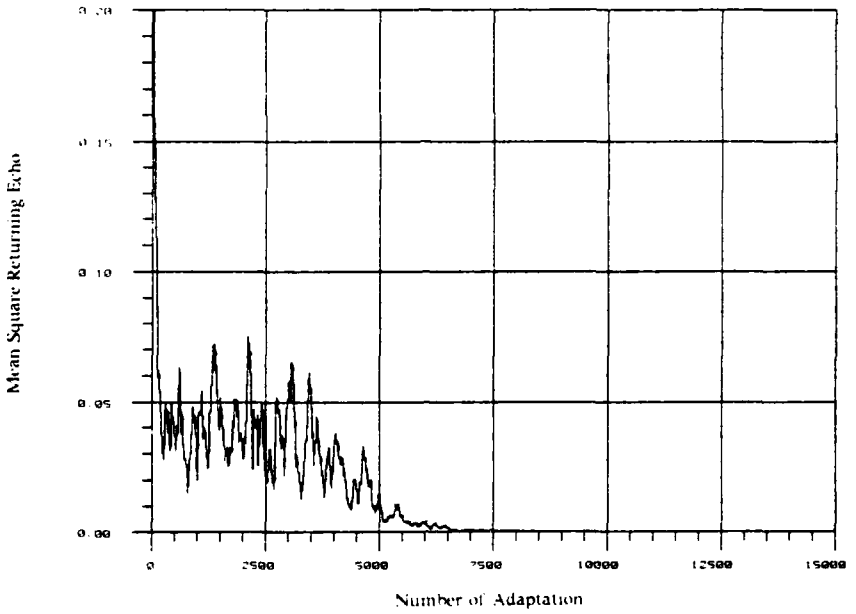


Figure 27 Mean square returning echo for white noise input

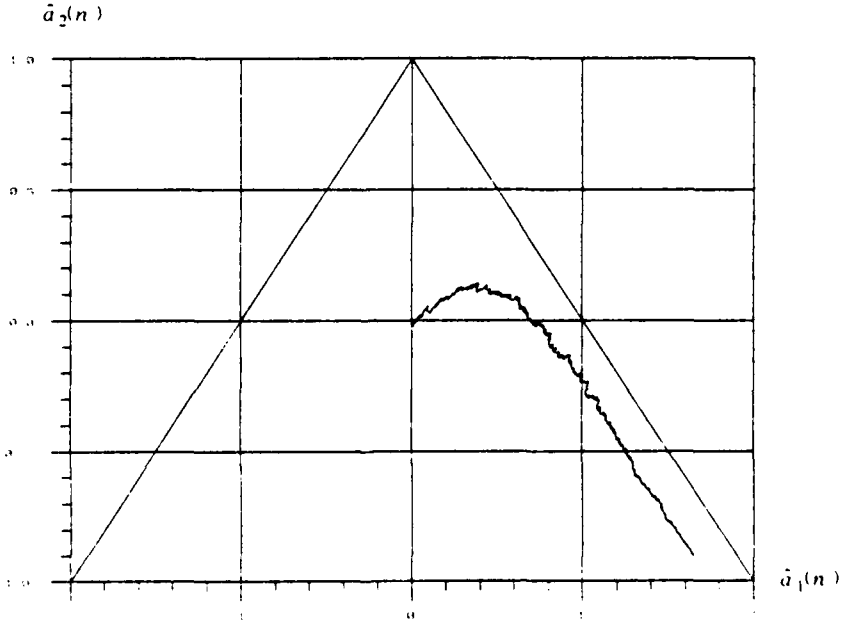


Figure 28 Convergence path of the denominator for white noise input

number of adaptation decreases as the input signal changes from the speech signal to the white noise. However, parameter convergence is achieved for all these three cases, although Figure 24 exhibits more irregularity for the speech signal input case. Rigorously speaking, speech signals are not stationary. Hence, they violate one of the basic assumptions for the input signal in the convergence proof (Section 2.3). In practice, however, the generalization of white noise to speech signals when one proceeds from theory to practice is widely accepted [13] - [19]. The above computer simulations also show that this generalization is not at all unreasonable. In other words, the conditions on the input signals for the convergence might be relaxed.

Next, the experiment for the same adaptive IIR echo canceler with these three inputs is repeated and compared with an adaptive FIR echo canceler. The number of weights in the FIR canceler is 32. They are all set to zero at the beginning of the adaptation. The gain is also normalized to $\frac{\tau}{\sum_{i=0}^{31} x(n-i)^2}$ where τ (step size) is always chosen to be the same as that in the IIR canceler (0.04). The number of multiplications at each adaptation is roughly 32 for the FIR canceler, and 9 for the IIR canceler. The ratio is about 3.5. Two kinds of comparison between the FIR canceler and the IIR canceler should be kept in mind when viewing the simulation results: i) for the same amount of computation, a comparison of performance; and ii) for the same performance, a comparison of computation. Figures 29 and 30 show the results with the speech signal input. The mean square returning echo is shown in Figure 29, and the ERLE defined by (59) (note that $v(n) \equiv 0$) is shown in Figure 30. Note that due to the large number of adaptations required, the input sentence is repeated for as many times as needed. It is seen from Figure 30 that the FIR canceler achieves a steady state ERLE (20 - 50 dB) after 37,500 adaptations whereas the IIR canceler needs more adaptation

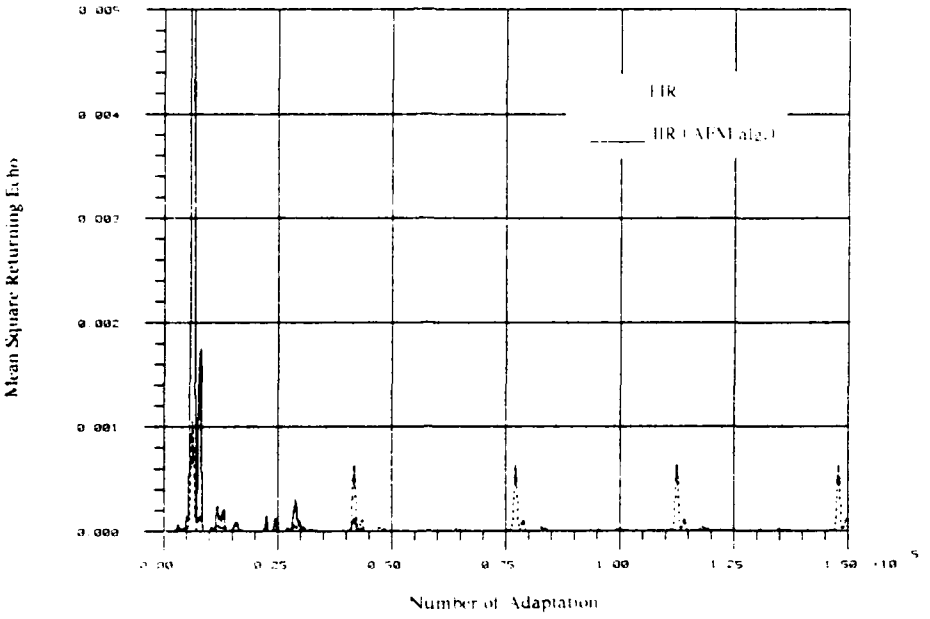


Figure 29 Mean square returning echoes for speech input, compared with an FIR canceler

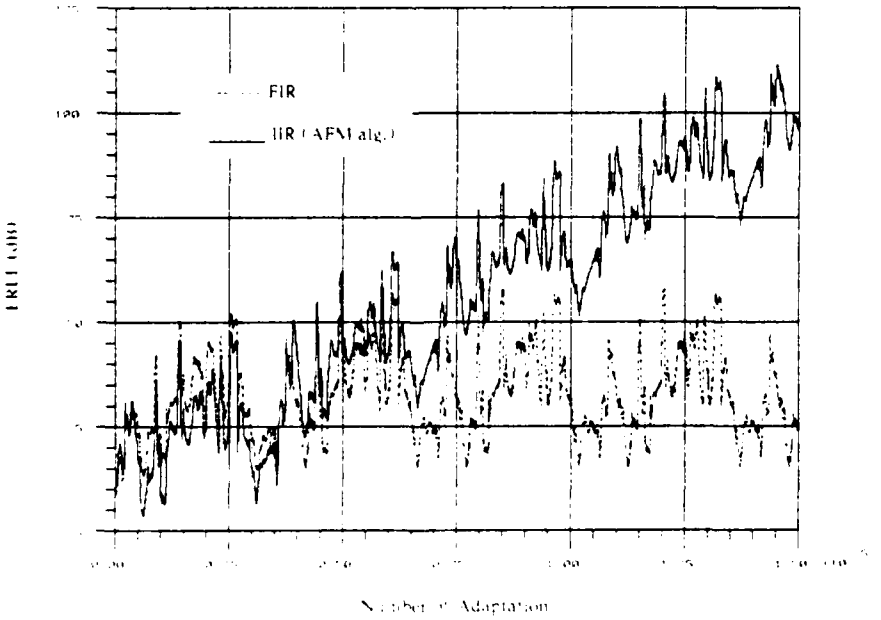


Figure 30 ERLE's for speech input, compared with an FIR canceler

and can achieve higher ERLE. For the same amount of computation as for the FIR canceler with 37,500 adaptations, the IIR canceler can have about $3.5 \times 37{,}500 \approx 130$k adaptations where its ERLE is 75 - 110 dB. On the other hand, for the IIR canceler to achieve 20 - 50 dB ERLE, it only needs 40k adaptations, which is much less than 130k. Thus, the IIR canceler performs much better than the FIR canceler in these senses. The potential for the IIR canceler is even more exciting: at 150k adaptations in Figure 30, the ERLE reaches the average of 100 dB and is still increasing. Figures 31 and 32 show the same comparison for the babble input, and Figures 33 and 34 are for the white noise input. The IIR canceler outperformed the FIR canceler in the above sense for all these three inputs. However, the price is slower convergence. It is especially clear from Figures 33 and 34 for the white noise input. In Figure 34, the FIR canceler quickly converges to its steady state value 22 dB after 2500 adaptations, whereas for the IIR canceler it takes 6250 adaptations, although the amount of computation at this point for the IIR canceler is only about 70 percent of that of the FIR canceler at 2500 adaptations.

The potential that the IIR canceler can achieve high ERLE is further investigated for the white noise input. Figure 35 shows an overall view of Figure 34, where it is seen that after 40k adaptations the ERLE of the IIR canceler settles at 110 dB level. It is apparent that the IIR canceler outperforms the FIR canceler tremendously in this sense.

The effect of τ in the convergence of the ERLE for the IIR canceler is shown in Figure 36 for τ=0.02, 0.04, and 0.08, respectively. The curve in the middle (τ=0.04) corresponds to that in Figure 35. It is seen that the slope of the curve for τ=0.08 is twice as much as the slope for the curve τ=0.04, which is in turn twice as much as that for τ=0.02. This is simply because $t = n\tau$. In other words, the convergence speed is

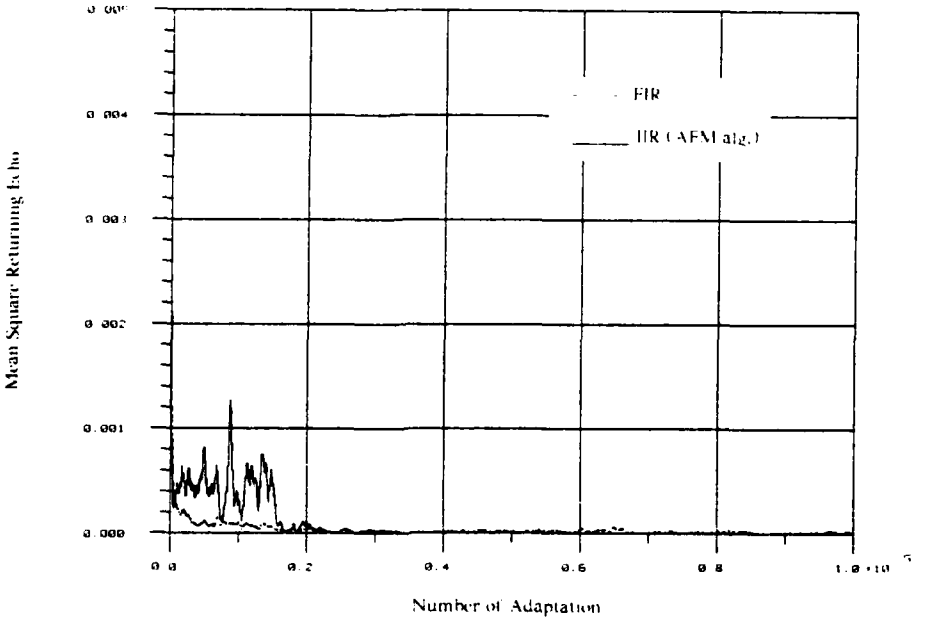


Figure 31 Mean square returning echoes for babble input, compared with an FIR canceler

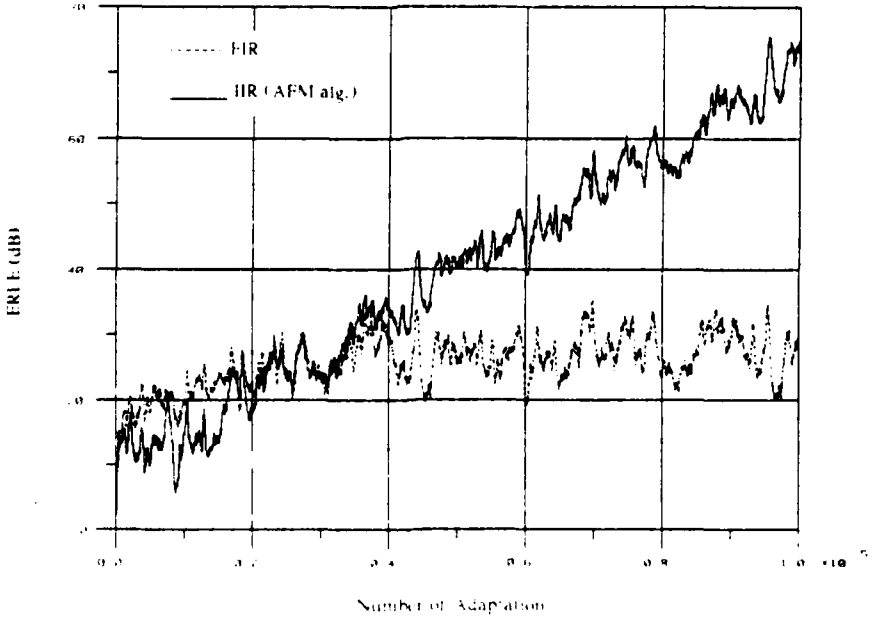


Figure 32 ERLE's for babble input, compared with an FIR canceler

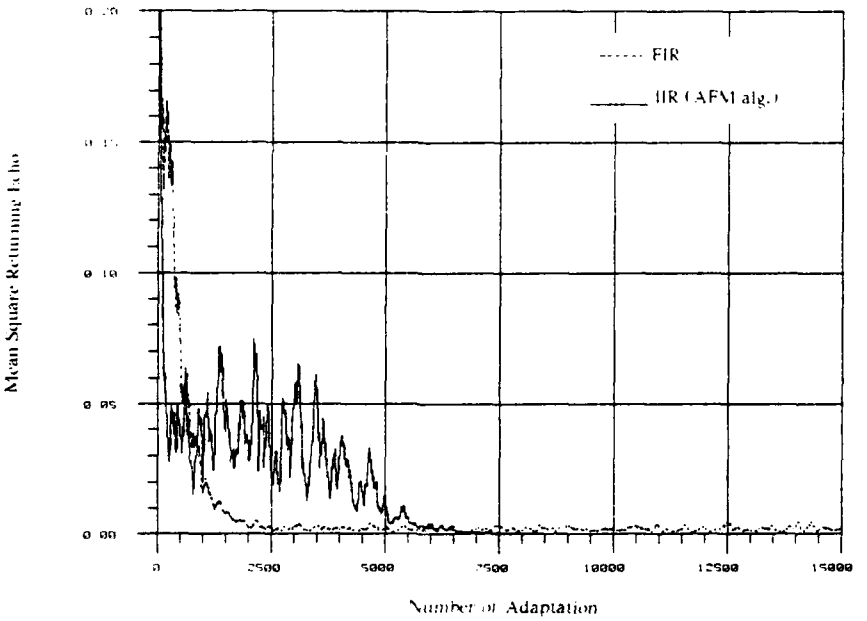


Figure 33 Mean square returning echoes for white noise input, compared with an FIR canceler

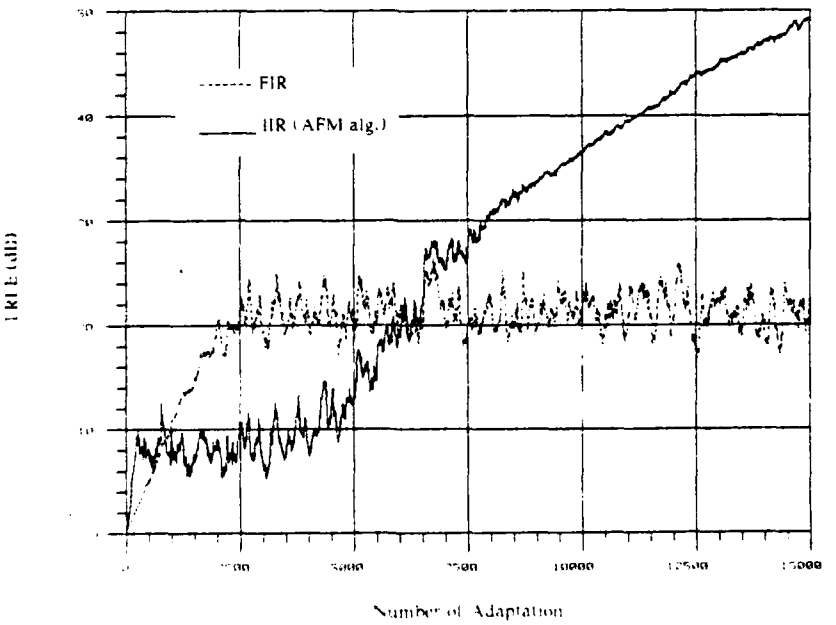


Figure 34 ERLE's for white noise input, compared with an FIR canceler

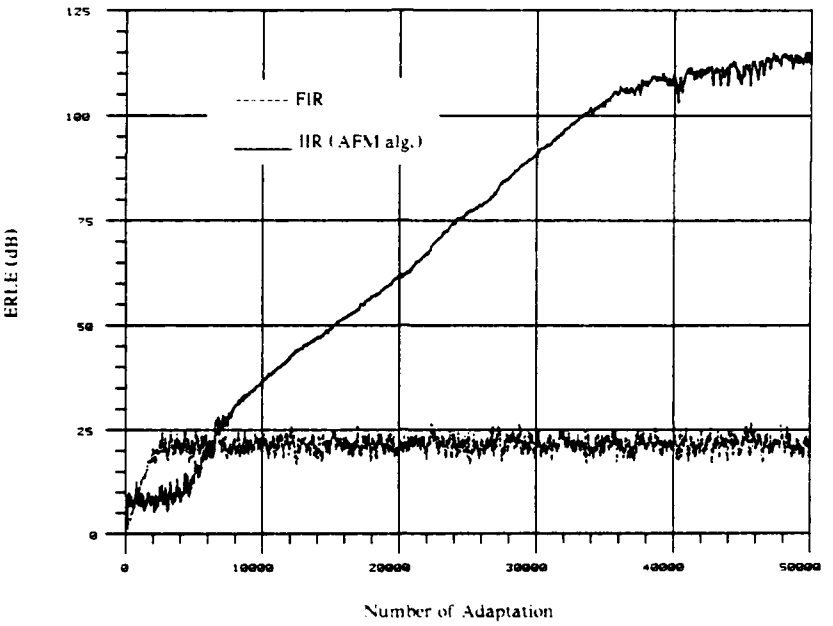


Figure 35 An overall view of Figure 34

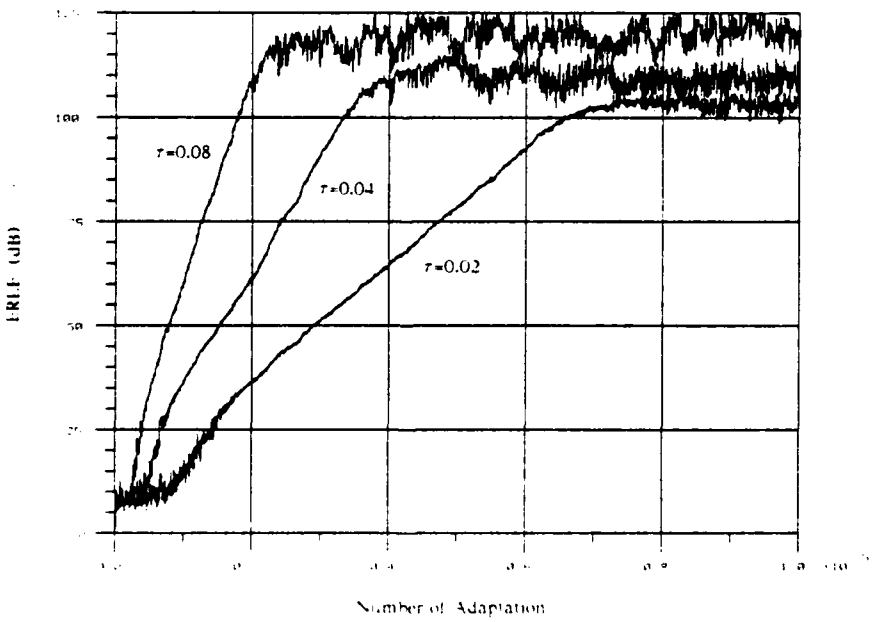


Figure 36 ERLE's for different values of τ, white noise input

proportional to τ. The variance of the ERLE at the convergence point, as can be seen from the amount of variation for the leveled off part of the three curves on the top of Figure 36, also increases as τ does, as expected. What is surprising in this figure is that the converged mean value of the ERLE also depends on τ, and increases slightly as τ does. At this point its cause is not clear yet.

The AFM algorithm is also compared with SHARF algorithm (7) for IIR cancelers. In implementing SHARF algorithm, the error smoothing filter is selected as $c_1=-0.8$ and $c_2=0$ so that the SPR requirement in (8) is satisfied. In order to achieve comparable results, the adaptive gain in SHARF algorithm is also normalized to $\frac{\tau}{\sum_{i=1}^{\hat{n}_a}\hat{y}(n-i)^2+\sum_{j=0}^{\hat{n}_b}x(n-j)^2}$ with $\tau=0.04$ which is the same as that used in the AFM algorithm. With the speech signal as the input, the results are shown in Figures 37 and 38. Differing from the AFM algorithm, SHARF starts to show significant returning echo reduction fairly late, but then it reaches the steady state very quickly. This is because SHARF uses a different convergence strategy than AFM. The convergence path of the adaptive coefficients for SHARF is very different from that of AFM. Note in Figure 38 that the steady state ERLE levels for the two IIR canceler algorithms are almost the same.

All the above experiments are conducted assuming $v(n)\equiv 0$. In reality, $\{v(n)\}$ is generally non-zero, and can be modeled as a white noise sequence. An experiment similar to that shown in Figures 29 and 30 is performed with $\{v(n)\}$ being a zero mean white Gaussian noise sequence with variance 0.0001. The same IIR canceler and FIR canceler are used. Due to the disturbance $v(n)$, the (normalized) adaptive gain must be much smaller than before ($\tau=0.0016$) to achieve satisfactory convergence and to ensure the stability of the IIR canceler. In order to avoid an excessive number of adaptations

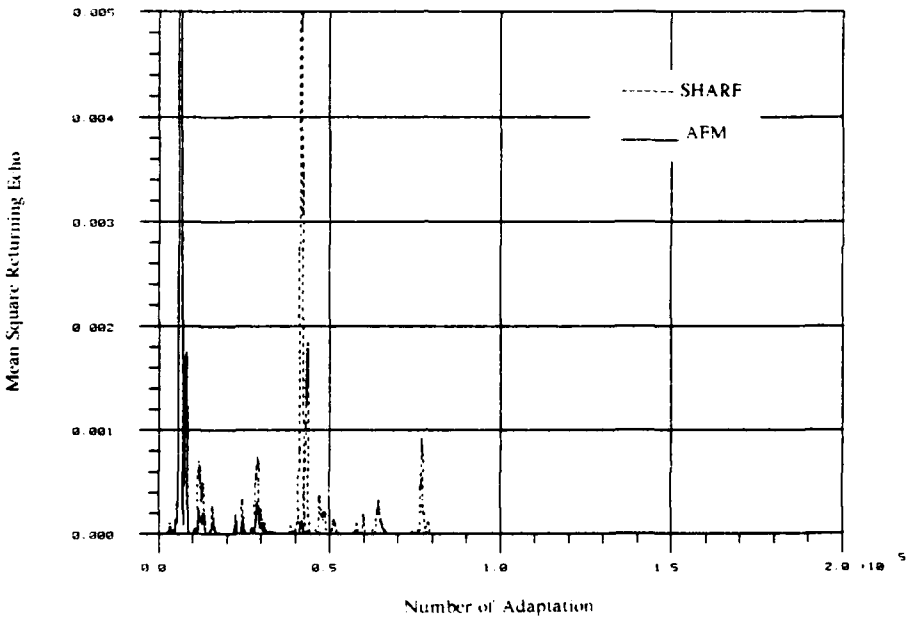


Figure 37 Mean square returning echoes for speech input, compared with SHARF

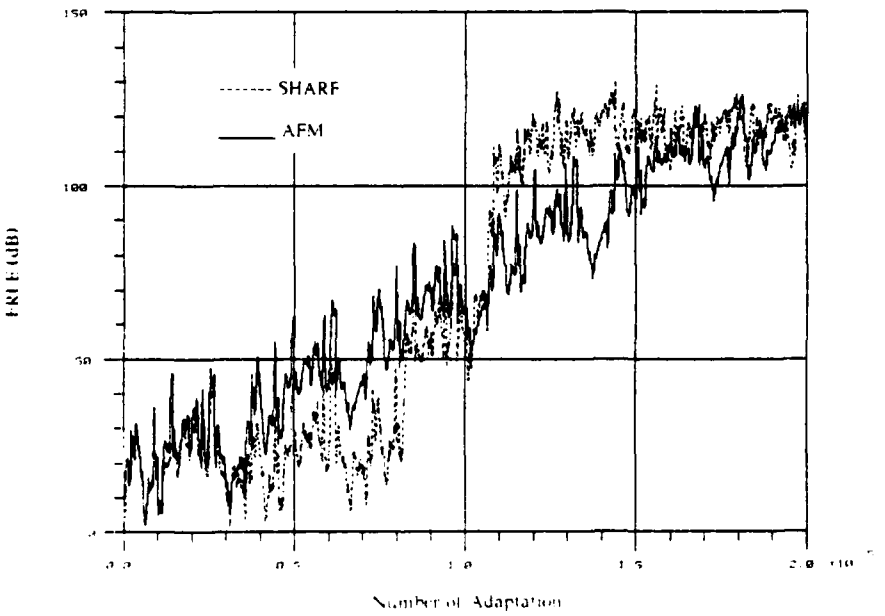


Figure 38 ERLE's for speech input, compared with SHARF

for such a small adaptive gain, the IIR canceler is initialized near the convergence point. It is chosen that $\hat{a}_1(0)=1.6$, $\hat{a}_2(0)=-0.8$, $\hat{b}_0(0)=0.5$, $\hat{b}_1(0)=-0.8$, and $\hat{b}_2(0)=0.5$. The FIR canceler is initialized with the values of the first 32 samples of the impulse response of the initial IIR canceler, $\hat{B}(0,z^{-1})/\hat{A}(0,z^{-1})$, which is

$$h(n)=0.625\delta(n)-2\,\mathrm{Re}\{(0.0625+j\,0.125)(0.8+j\,0.4)^n\};\qquad n=0,1,\cdots,31. \quad (60)$$

This initialization is also reasonable in a practical sense. Usually an echo canceler is used for a long time after installation; hence, every time it is turned on it always starts adaptation from some nominal value near the optimum rather than from zero as it is first installed.

The results of this experiment are shown in Figures 39 and 40. Note that for the definition of ERLE as in (59), when the signal part in $y(n)$ becomes small, as a speech signal often does, both $y(n)$ and $e(n)$ are dominated by $v(n)$ and the ERLE becomes zero. Thus, the ERLE curve constantly oscillates between zero and its functional value. Our experience indicates that the situation can be so bad that one cannot obtain any information from the ERLE curve thus defined, and hence, it essentially becomes useless. A modified definition used in [18] can overcome this difficulty:

$$ERLE=10\log\frac{E\{[y(n)-v(n)]^2\}}{E\{[e(n)-v(n)]^2\}}. \quad (61)$$

For $v(n)\equiv 0$, (61) coincides with (59). Note that this definition is not practical since $v(n)$ is not measurable. However, it does give some indication about how well an echo canceler is functioning in the presence of $v(n)$, especially for real speech input. The ERLE curves in Figure 40 are the results of the experiment using definition (61). It can be seen from these two figures that the IIR canceler still outperforms the FIR canceler while both are degraded. However, the improvement of the IIR canceler over the FIR is much less than the noise-free case, because the noise causes the adaptive coefficients to

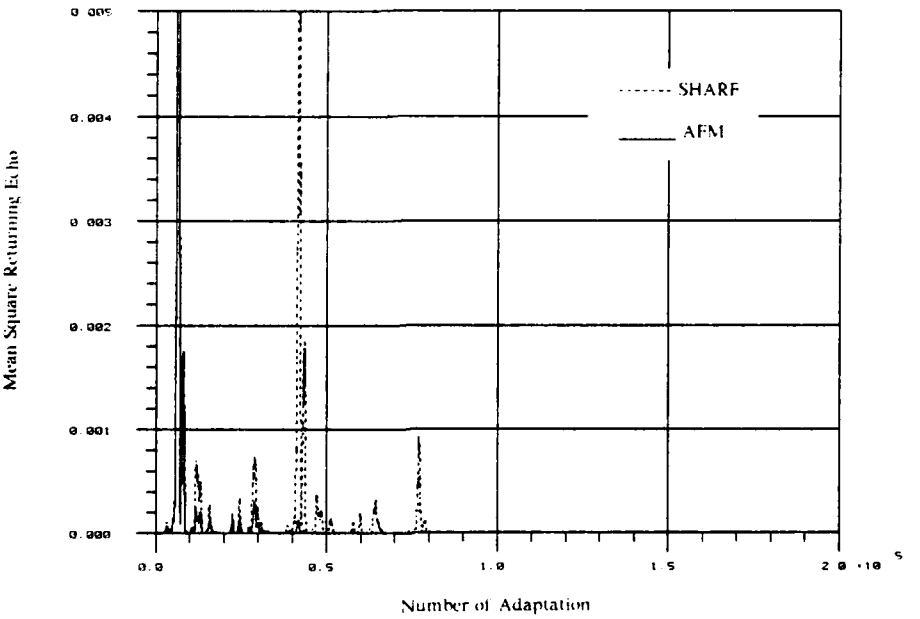


Figure 37 Mean square returning echoes for speech input, compared with SHARF

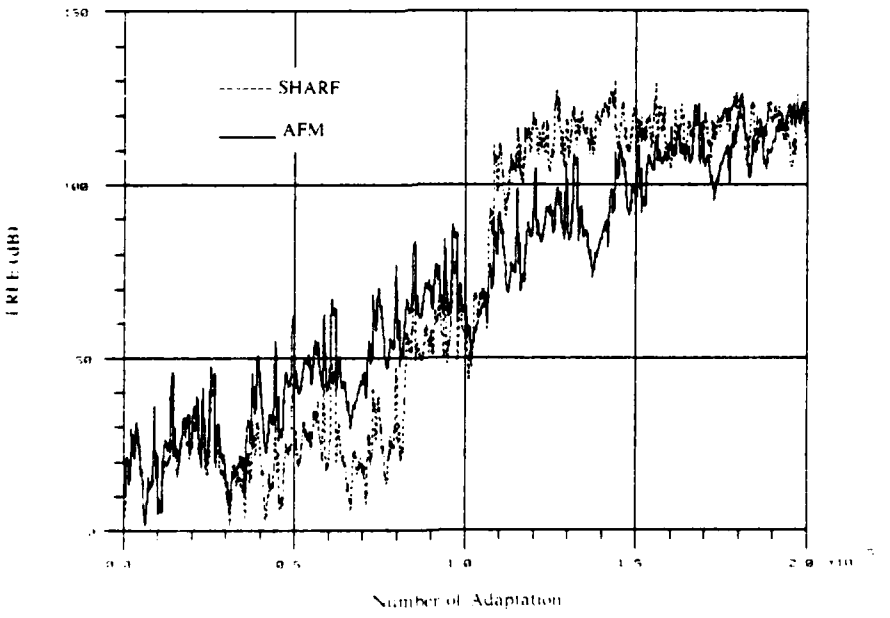


Figure 38 ERLE's for speech input, compared with SHARF

wander around the true value with a much larger variance.

3.2.2 Reduced Order Case

To construct a reduced order situation, the echo path transfer function of the last subsection (sufficient order case) is modified by adding a pole at 0.7 and a zero at 1.4286 so that it still has an all-pass characteristic (see Figures 41 and 42). The gain is 0.3 so that the loss is about 6.5 dB. The same adaptive IIR canceler and FIR canceler are used with the three inputs for $v(n) \equiv 0$. The constant τ in the normalized gain of the cancelers is adjusted to have the value 0.01. All the adaptive coefficients are initialized with zero value.

Figures 43 and 44 show a similar comparison between the IIR canceler and the FIR canceler as before. It may seem *surprising that the* IIR canceler not only performs poorer than the FIR, it also shows nearly no reduction in the mean square returning echo (Figure 43) and no increase in ERLE (Figure 44). This is probably because of i) possible shallow holes of the error surface as a (usually complicated) function of the five adaptive coefficients, and of ii) the non-stationarity of the speech signal. It is obvious that the global minimum value of the error surface for the IIR canceler decreases as the number of the adaptive coefficients increases. It reaches zero when a sufficient order situation is obtained. Similarly, the only minimum value of the *quadratic* error surface for the FIR canceler also decreases as the number of weights increases. However, it will never reach zero if the unknown dynamic plant has a rational transfer function. What happens here is probably that the global minimum of the error surface for the adaptive IIR canceler is larger than the minimum of the *quadratic* error surface for the adaptive FIR canceler. It should not be surprising to see further deduction in the mean square returning echo (eventually to zero) using the IIR

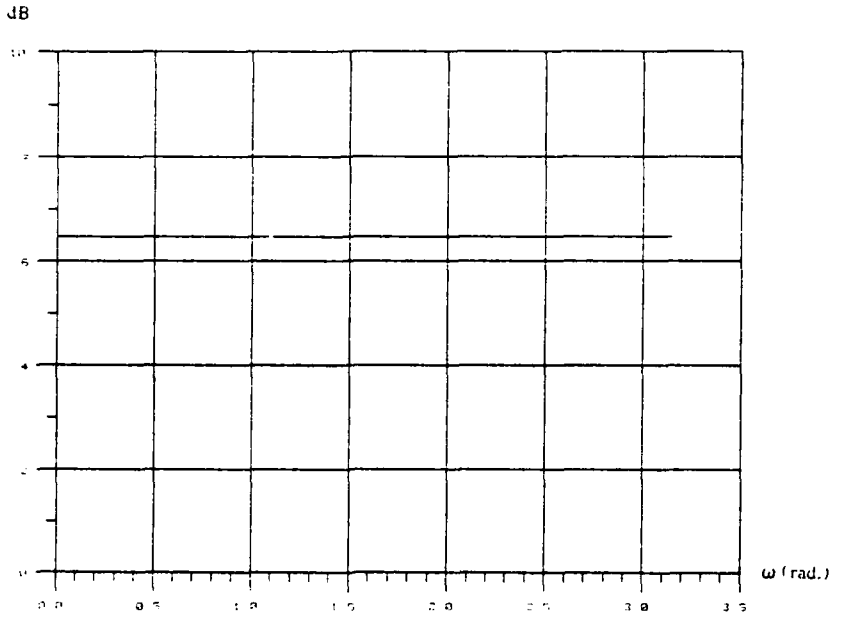


Figure 41 Magnitude (loss) of the 3rd order echo path transfer function

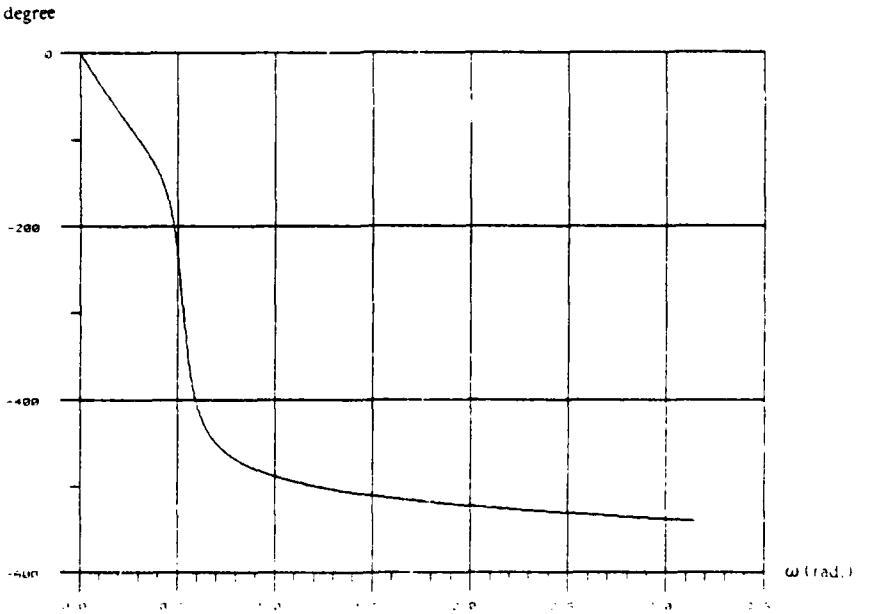


Figure 42 Phase of the 3rd order echo path transfer function

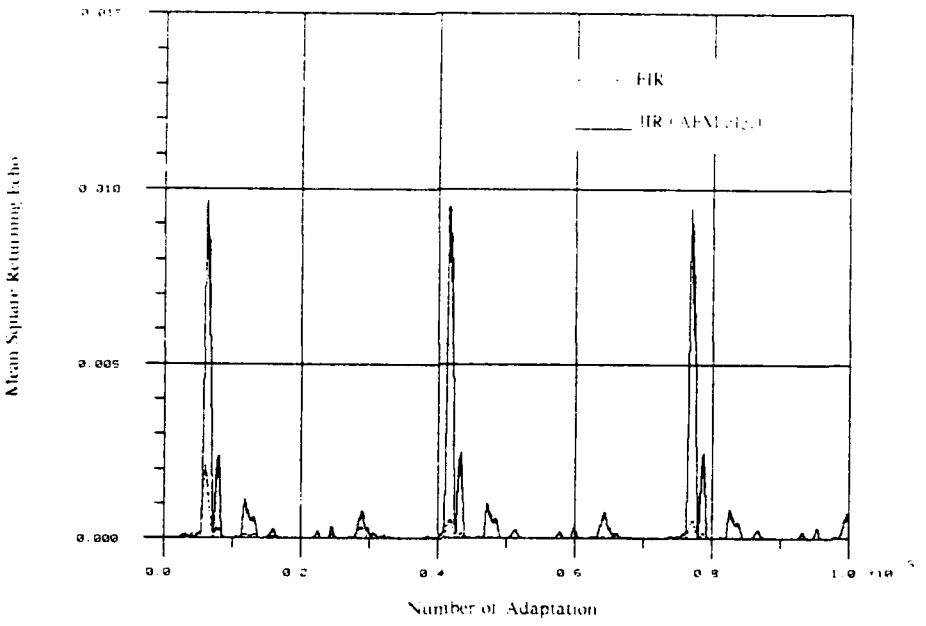


Figure 43 Mean square returning echoes for speech input, compared with an FIR canceler, reduced order

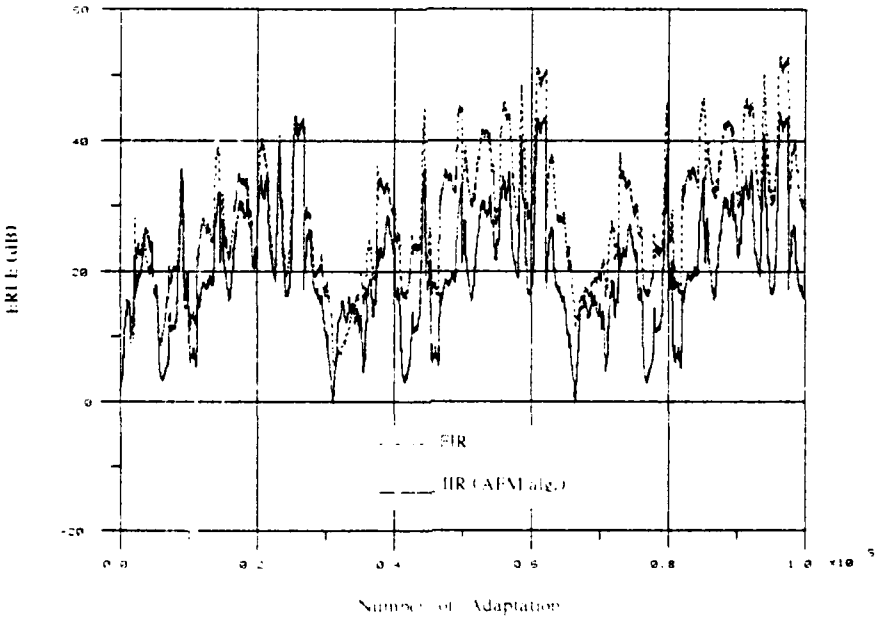


Figure 44 ERLE's for speech input, compared with an FIR canceler, reduced order

canceler if the number of adaptive coefficients increases.

An overall view of the above situation for the IIR canceler can be seen in Figures 45 and 46 for the returning echo and the convergence path of the coefficients in the denominator. The adaptive coefficients do converge to some point as seen in Figure 46. It is interesting to see the repeating pattern of the convergence path as the input sentence is repeated. Even at the convergence point, the adaptive coefficients circle along a "P" shape, resulting in the large error at a certain portion of the input signal (Figure 45). This is due to the peculiarity (e.g., non-stationarity) of the speech signal, which is not obviously shown in the sufficient order case before. "Better" input signals: the babble signal and the white noise sequence are used for the same experiment, and the results are shown in Figures 47 - 50. For the babble input (Figures 47 and 48), decrease of the mean square returning echo for some portion of the input signal is obvious while increase for some other portion still exists. The convergence of the adaptive parameters is seen better than that with the speech signal input in the sense that the variance is much smaller. White noise is perhaps the most ideal input (Figures 49 and 50). Decrease of the mean square returning echo and convergence to approximately the same point is observed. It should be noted that the mean square returning echo reaches a steady value at only 10k adaptations (two percent of the total number of adaptations), whereas it took almost 500k adaptations to achieve parameter convergence.

The above IIR canceler (using AFM algorithm) is also compared with the one using Stearns' algorithm (6) for this reduced order case. The gain in (6) is also normalized to $\frac{\tau}{\sum_{i=1}^{n_a}\hat{y}'(n-i)^2+\sum_{j=0}^{n_b}x'(n-j)^2}$ with τ=0.01. From the results in Figures 51 and 52 it is seen that the AFM algorithm performs slightly better than Stearns' algorithm. Since Stearns' algorithm is a gradient algorithm, this verifies that the AFM algorithm has not

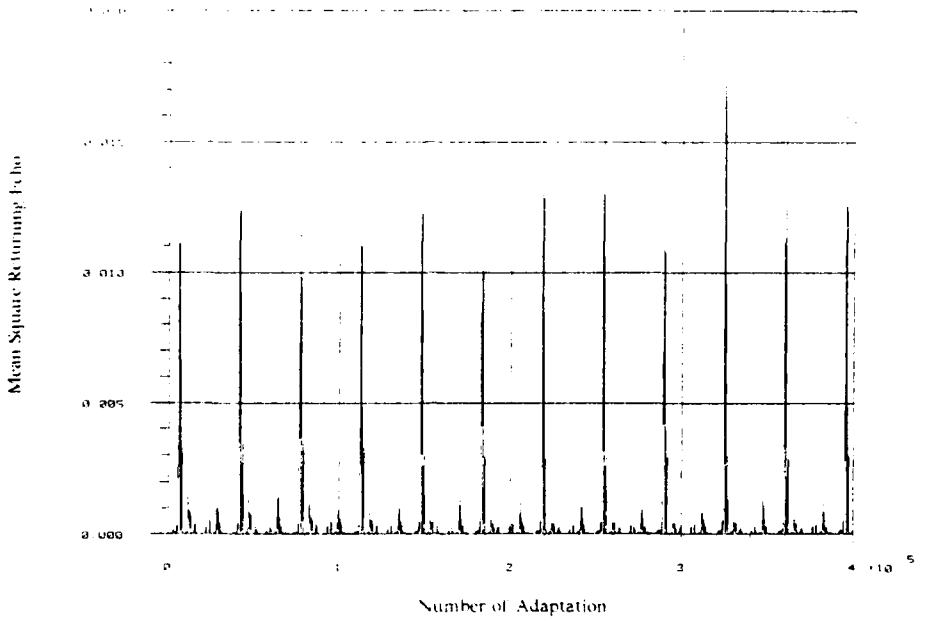


Figure 45 Mean square returning echo for speech input, reduced order

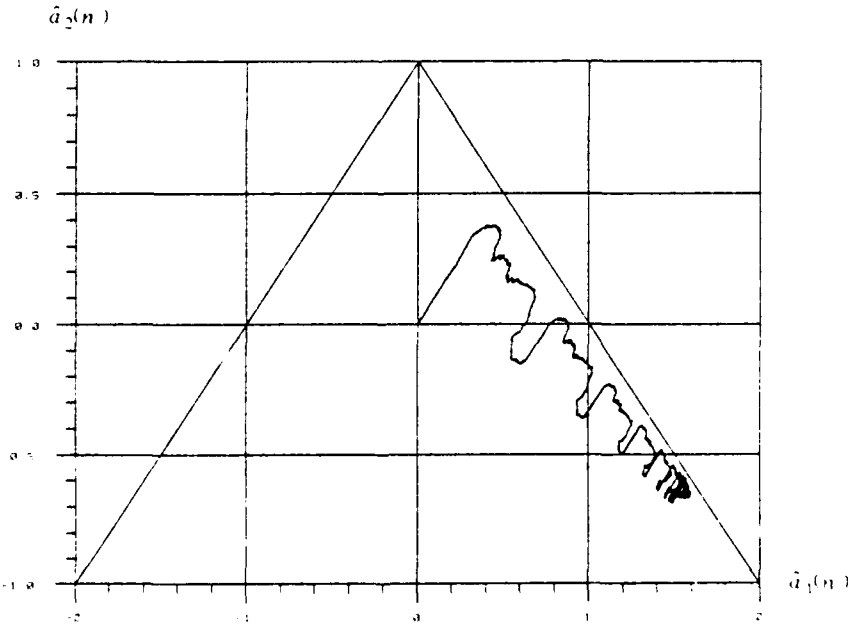


Figure 46 Convergence path of the denominator for speech input, reduced order

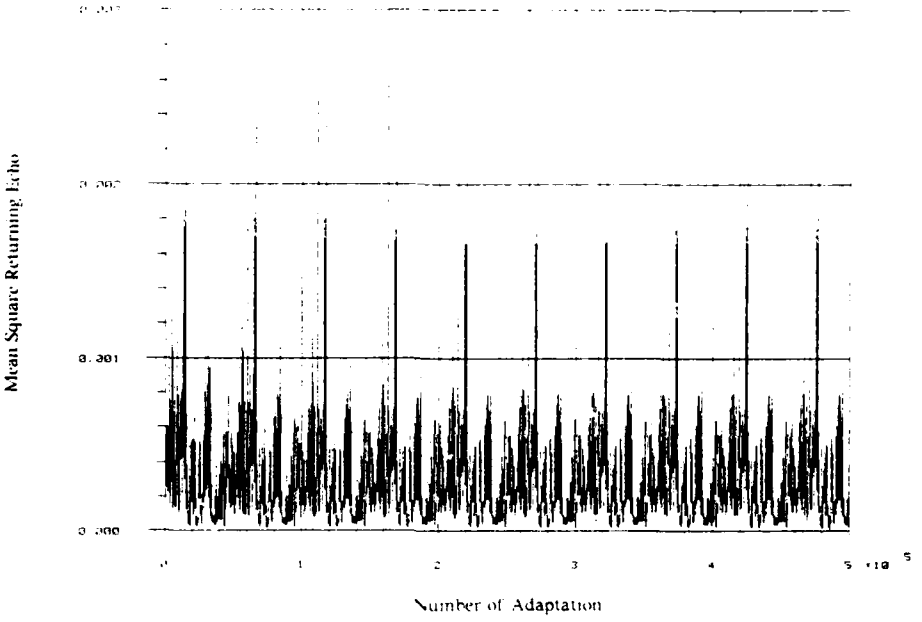


Figure 47 Mean square returning echo for babble input, reduced order

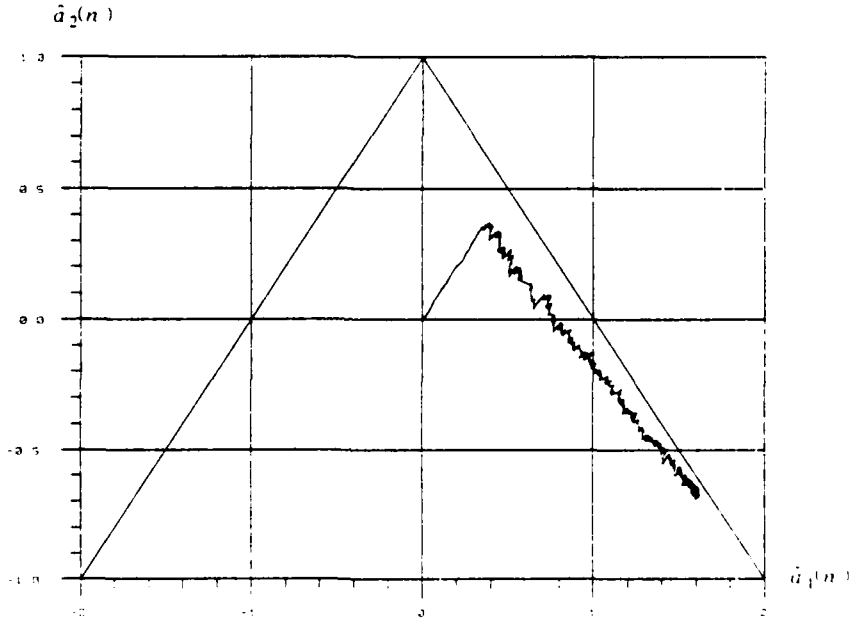
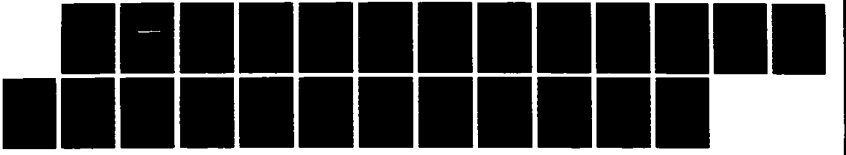
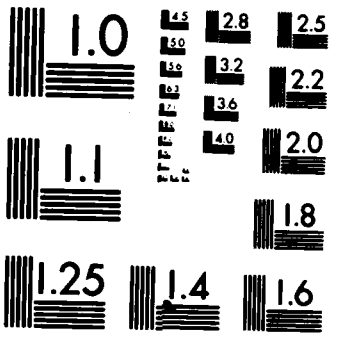


Figure 48 Convergence path of the denominator for babble input, reduced order

AD-A171 894 NEW ADAPTIVE IIR FILTERING ALGORITHMS(U) ILLINOIS UNIV 2/2
AT URBANA COORDINATED SCIENCE LAB H FAN AUG 86
UILU-ENG-86-2224 N00014-84-C-0149

UNCLASSIFIED F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

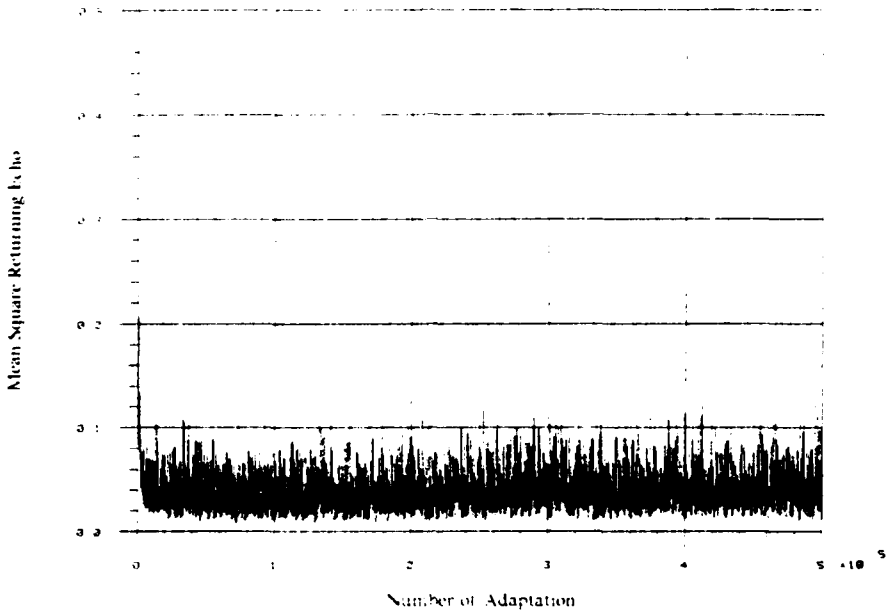


Figure 49 Mean square returning echo for white noise input, reduced order

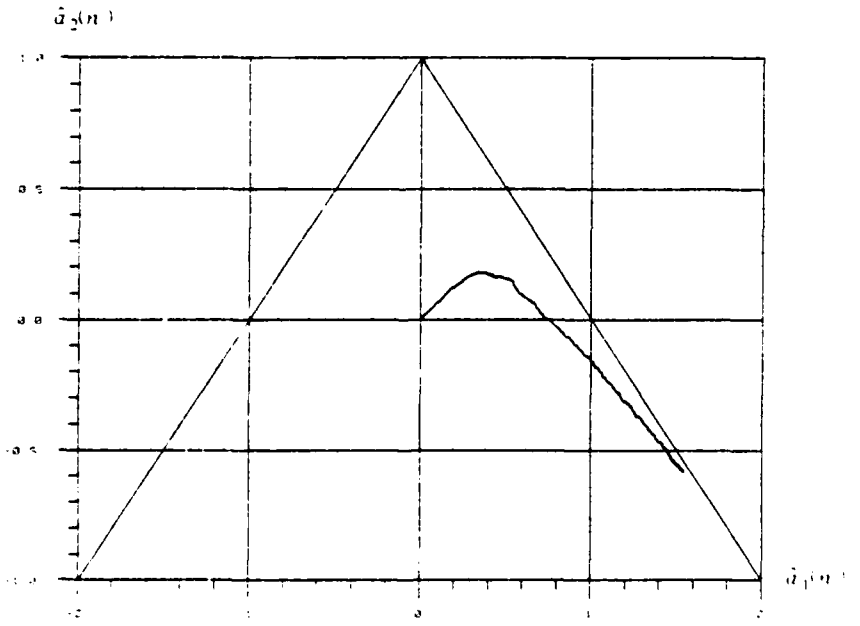


Figure 50 Convergence path of the denominator for white noise input, reduced order

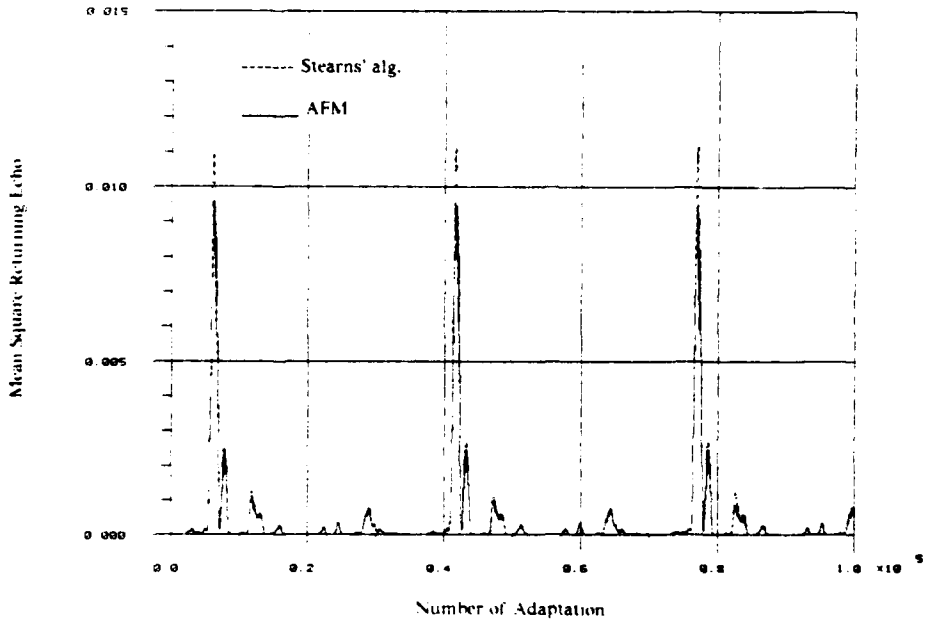


Figure 51 Mean square returning echoes for speech input, compared with Stearns' algorithm, reduced order

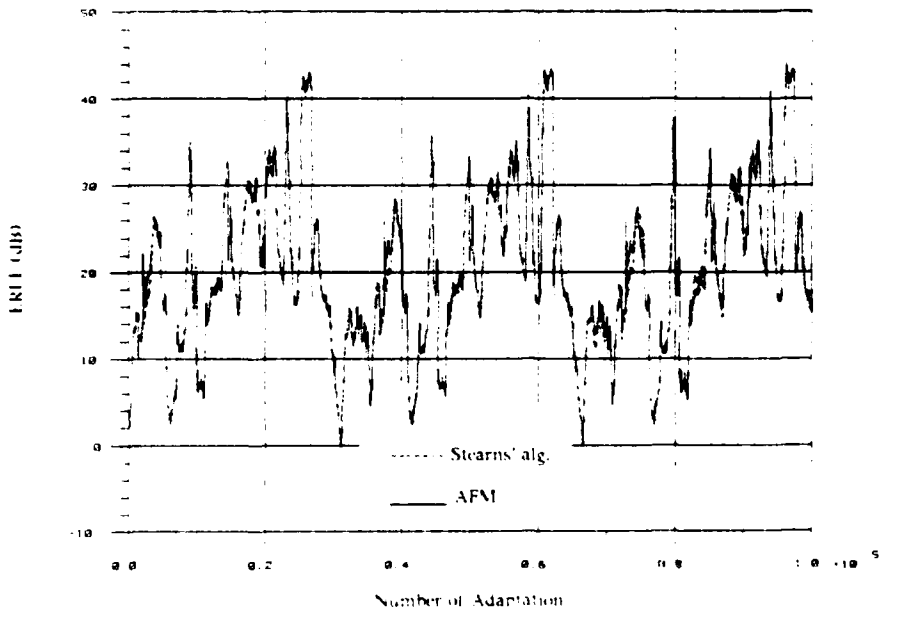


Figure 52 ERLE's for speech input, compared with Stearns' algorithm, reduced order

missed other local minimum points that are lower than the one that AFM is aiming at.

Finally, the AFM canceler is compared with the SHARF canceler used in the last subsection with τ changed to 0.01. Note that the SPR requirement (8) is no longer satisfied with c_1=-0.8 and zero otherwise. However, it can be checked that the SPR requirement is violated for only a small amount on a small portion of the unit circle. Based on the argument that the SPR requirement might be overly restrictive [59], the SHARF canceler is implemented the same as before without any other modification. Figures 53 and 54 show the results, from which SHARF is seen to perform much worse than the AFM algorithm. This confirms the observation of SHARF's misbehavior in the reduced order case [62], [64].

From the above rich collection of examples it is clearly seen that, in general, adaptive IIR cancelers may not perform as well as adaptive FIR cancelers in reduced order case. However, if one selects the order correctly, e.g., if a 3rd order adaptive canceler is used in this subsection, it is obvious that the same results as those in the last subsection for sufficient order case may be expected. In other words, the 3rd order IIR canceler using the AFM algorithm would outperform the FIR canceler having 32 (or more) taps, although more adaptations may be needed. No more such examples shall be included here.

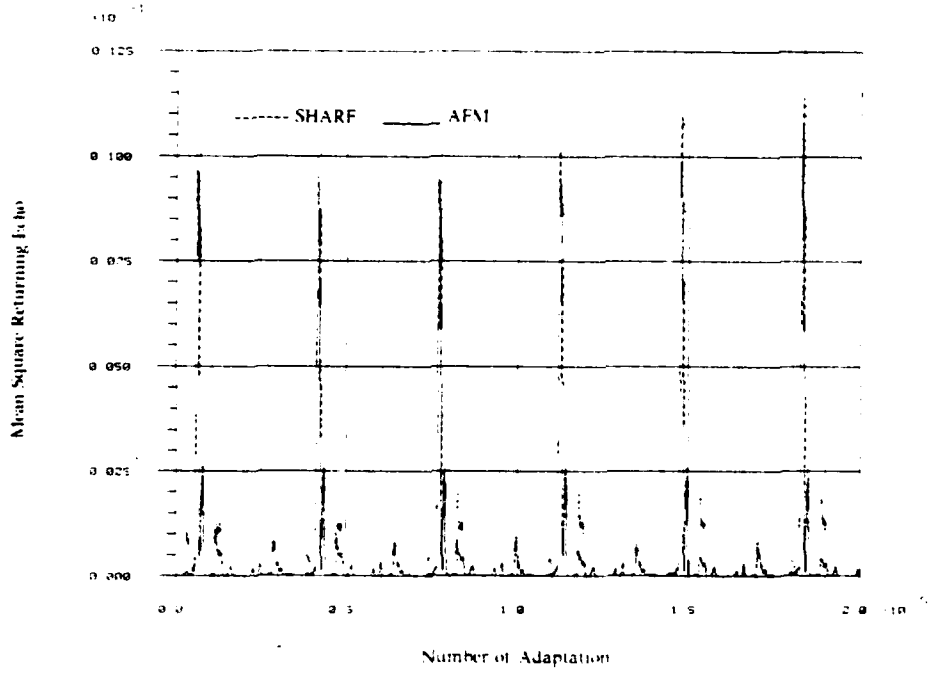


Figure 53 Mean square returning echoes for speech input, compared with SHARF, reduced order

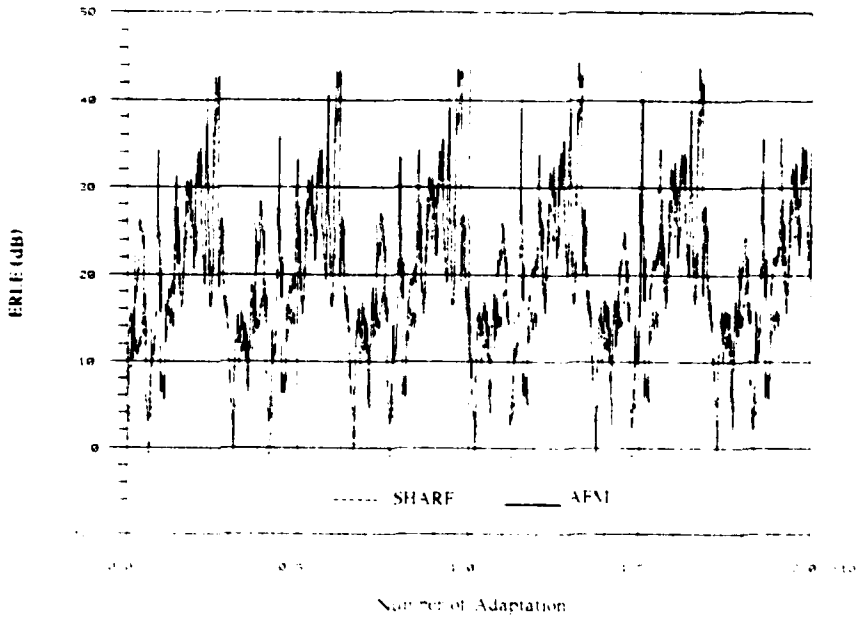


Figure 54 ERLE's for speech input, compared with SHARF, reduced order

4. CONCLUSION

A family of new adaptive IIR filtering algorithms, SIM algorithm, AFM algorithm, and IF algorithm, is proposed based on the Steiglitz-McBride identification scheme. It is proved using a theorem of wide-sense convergence in probability that the IF algorithm converges to the solution of an associated ODE under fairly general conditions. It is further proved, imposing more conditions, that the filter coefficients converge to the true system parameter values in sufficient order case. The AFM algorithm, which is shown to be a close approximation of the IF algorithm, is implemented to verify the above theoretical results. For reduced order case, a number of computer simulations using the AFM algorithm exhibits global convergence regardless of local minima. The proposed algorithms are also applied to the problem of echo cancellation. A large variety of simulations using various input signals is performed and compared with adaptive FIR echo cancelers and other existing adaptive IIR algorithms. The results are favorable.

It is shown in Section 2.2 that the family of algorithms is derived from traditional equation error approach. With pre-filtering, the inherent flaw of the equation error method, parameter bias in the presence of a disturbance, is removed to the extent that only a colored disturbance would cause parameter bias. The effect of this bias in relation to the input and the unknown plant is studied. For a white disturbance, the algorithm is shown to converge exponentially.

It should be pointed out that the problem of the adaptive IIR filtering in the reduced order case is quite difficult. Some researchers approached this problem by finding error bounds so that the filter can be guaranteed at least not to "blow up" [86], [87]. The global convergence of the proposed algorithms in Section 2.4 is unique and interesting. Although this observation only gives a limited view, it deserves the following conjecture. For reduced order cases, if all conditions in Theorem 1 and

Theorem 2 of Section 2.3 are met, then the IF algorithm converges to the *best fit*, i.e., the global minimum point of the mean square error surface, if such a minimum exists. In practice, however, the effectiveness of using adaptive IIR filters as opposed to adaptive FIR filters depends largely on the correct choice of the filter orders. Even with the global convergence, an adaptive IIR filter may still be inferior to an adaptive FIR filter in the reduced order case, as seen from the application examples in Section 3.2.2.

One major purpose of having constant gain for adaptive filtering is to track time-varying systems. However, due to the complexity of time-varying situations, not much analysis has been done [72], [88]. More powerful convergence theorems (e.g., [74]) may be needed, and stronger conditions may have to be imposed (e.g., [89] to allow $S \rightarrow \infty$ in (29) and (30)). Having dealt with only the stationary situation and a specific area of application, this dissertation can be a necessary pre-study for aforementioned more complicated cases, and be a demonstration of potentially wider application of recent weak convergence results [71] - [74] to various adaptive filtering algorithms [49] - [51], [59] - [67]. The proposed algorithm might be further modified to estimate the disturbance so that the bias in the presence of colored disturbances can be removed. The problem of stability monitoring needs to be studied further. Other areas of application are also possible. In short, the adaptive IIR filtering is a young and prosperous research area. The proposition of the family of new algorithms only opens an additional window which will hopefully bring in some fresher air.

APPENDIX

PROOF OF THE LEMMAS

A) Proof of Lemma 1:

From (34), it follows that

$$Y(n,\hat{\theta}_1)=F(\hat{\theta}_1)^n Y(0)+\sum_{i=0}^{n-1}F(\hat{\theta}_1)^i G(\hat{\theta}_1)U(n-i)$$

$$Y(n,\hat{\theta}_2)=F(\hat{\theta}_2)^n Y(0)+\sum_{i=0}^{n-1}F(\hat{\theta}_2)^i G(\hat{\theta}_2)U(n-i)$$

Thus,

$$\begin{aligned}\| Y(n,\hat{\theta}_1)-Y(n,\hat{\theta}_2)\| &= \| [F(\hat{\theta}_1)^n-F(\hat{\theta}_2)^n]Y(0)\\ &\quad +\sum_{i=0}^{n-1}[F(\hat{\theta}_1)^i G(\hat{\theta}_1)-F(\hat{\theta}_2)^i G(\hat{\theta}_2)]U(n-i)\|\\ &\leq \| F(\hat{\theta}_1)^n-F(\hat{\theta}_2)^n \| \, \| Y(0)\|\\ &\quad +\sum_{i=0}^{n-1}\| F(\hat{\theta}_1)^i G(\hat{\theta}_1)-F(\hat{\theta}_2)^i G(\hat{\theta}_2)\| \, \| U(n-i)\| \qquad \text{(A1)}\end{aligned}$$

We now estimate $\| F(\hat{\theta}_1)^n-F(\hat{\theta}_2)^n \|$ and $\sum_{i=0}^{n-1}\| F(\hat{\theta}_1)^i G(\hat{\theta}_1)-F(\hat{\theta}_2)^i G(\hat{\theta}_2)\| \cdot \| U(n-i)\|$ respectively.

Using the identity $A_1^n-A_2^n=A_1^n-A_1^{n-1}A_2+A_1^{n-1}A_2-A_1^{n-2}A_2^2+\cdots+A_1A_2^{n-1}-A_2^n$, the first estimate is

$$\| F(\hat{\theta}_1)^n-F(\hat{\theta}_2)^n \| = \| \sum_{j=0}^{n-1}F(\hat{\theta}_1)^{n-1-j}[F(\hat{\theta}_1)-F(\hat{\theta}_2)]F(\hat{\theta}_2)^j \|$$

$$\leq \sum_{j=0}^{n-1} \| F(\hat{\theta}_1)^{n-1-j} \| \, \| F(\hat{\theta}_1) - F(\hat{\theta}_2) \| \, \| F(\hat{\theta}_2)^j \|$$

$$\leq C_1 \| \hat{\theta}_1 - \hat{\theta}_2 \| \sum_{j=0}^{n-1} \| F(\hat{\theta}_1)^{n-1-j} \| \, \| F(\hat{\theta}_2)^j \| . \qquad \text{(A2)}$$

Let $\lambda_i(\hat{\theta})$ denote the i th eigenvalue of $F(\hat{\theta})$. By the assumption i), there exists $0<\lambda<1$ such that $|\lambda_i(\hat{\theta})|<\lambda$ for all i and all $\hat{\theta} \epsilon \hat{D}_c$. Hence, by [90, Theorem 5], there exists $C_3>0$ such that $\| F(\hat{\theta})^j \| \leq C_3 \lambda^j$.[4] So (A2) can be written as

$$\| F(\hat{\theta}_1)^n - F(\hat{\theta}_2)^n \| \leq C_1 C_3^2 \| \hat{\theta}_1 - \hat{\theta}_2 \| \sum_{j=0}^{n-1} \lambda^{n-1} = C_1 C_3^2 \| \hat{\theta}_1 - \hat{\theta}_2 \| n \lambda^{n-1}. \qquad \text{(A3)}$$

It is easily seen by comparing two adjacent terms that as n increases, the sequence $n\lambda^{n-1}$ reaches its maximum value (≥ 1) at $N = [\frac{\lambda}{1-\lambda}]$ where again $[\cdot]$ denotes the upper nearest integer, and then decreases monotonically to zero as $n \to \infty$. Hence, (A3) is bounded by

$$\| F(\hat{\theta}_1)^n - F(\hat{\theta}_2)^n \| \leq C_1 C_3^2 N \lambda^{N-1} \| \hat{\theta}_1 - \hat{\theta}_2 \| . \qquad \text{(A4)}$$

Next, estimate

$$\sum_{i=0}^{n-1} \| F(\hat{\theta}_1)^i G(\hat{\theta}_1) - F(\hat{\theta}_2)^i G(\hat{\theta}_2) \| \, \| U(n-i) \|$$

$$= \sum_{i=0}^{n-1} \| F(\hat{\theta}_1)^i G(\hat{\theta}_1) - F(\hat{\theta}_1)^i G(\hat{\theta}_2) + F(\hat{\theta}_1)^i G(\hat{\theta}_2) - F(\hat{\theta}_2)^i G(\hat{\theta}_2) \| \, \| U(n-i) \|$$

$$\leq \sum_{i=0}^{n-1} \| F(\hat{\theta}_1)^i \| \, \| G(\hat{\theta}_1) - G(\hat{\theta}_2) \| \, \| U(n-i) \|$$

---

[4] Although in [90, Theorem 5] the conclusion is $\| A^j \|_2 \leq C_3 \lambda^j$ where $\| A \|_2 = [\rho(A^T A)]^{1/2}$, it is readily seen by the equivalence of norms [78] that $\| A^j \| \leq C'_3 \| A^j \|_2 \leq C'_3 C_3 \lambda^j$.

$$+\sum_{i=0}^{n-1} \| F(\hat{\theta}_1)^i - F(\hat{\theta}_2)^i \| \, \| G(\hat{\theta}_2) \| \, \| U(n-i) \|$$

$$\leq C_2 C_3 \| \hat{\theta}_1 - \hat{\theta}_2 \| \sum_{i=0}^{n-1} \lambda^i \, \| U(n-i) \| + C_1 C_3^2 \, \| G(\hat{\theta}_2) \| \, \| \hat{\theta}_1 - \hat{\theta}_2 \| \sum_{i=0}^{n-1} i \lambda^{i-1} \| U(n-i) \|$$

where the last step is obtained using (A3). Now the total estimate is

$$\| Y(n, \hat{\theta}_1) - Y(n, \hat{\theta}_2) \| \leq \Big| C_1 C_3^2 N \lambda^{N-1} \| Y(0) \| + C_2 C_3 \sum_{i=0}^{n-1} \lambda^i \, \| U(n-i) \|$$

$$+ C_1 C_3^2 \, \| G(\hat{\theta}_2) \| \sum_{i=0}^{n-1} i \lambda^{i-1} \| U(n-i) \| \Big| \, \| \hat{\theta}_1 - \hat{\theta}_2 \|$$

$$= \Upsilon \, \| \hat{\theta}_1 - \hat{\theta}_2 \| .$$

It is obvious that $\Upsilon > 0$. Its mean is

$$E\{\Upsilon\} = C_1 C_3^2 N \lambda^{N-1} \| Y(0) \| + C_2 C_3 E\{ \| U(n-i) \| \} \sum_{i=0}^{n-1} \lambda^i$$

$$+ C_1 C_3^2 \, \| G(\hat{\theta}_2) \| E\{ \| U(n-i) \| \} \sum_{i=0}^{n-1} i \lambda^{i-1} .$$

Obviously $\sum_{i=0}^{n-1} \lambda^i = \frac{1-\lambda^n}{1-\lambda} \leq \frac{1}{1-\lambda} < \infty$. Now since $i \lambda^{i-1} = \frac{d}{d\lambda} \lambda^i$ and n is finite, it follows that

$$\sum_{i=0}^{n-1} i \lambda^{i-1} = \frac{d}{d\lambda} \sum_{i=0}^{n-1} \lambda^i = \frac{d}{d\lambda} \frac{1-\lambda^n}{1-\lambda} = \frac{1}{(1-\lambda)^2} - \frac{n \lambda^{n-1}(1-\lambda) + \lambda^n}{(1-\lambda)^2} \leq \frac{1}{(1-\lambda)^2} < \infty .$$

Thus by the assumptions it can be concluded that $E\{\Upsilon\} < \infty$.

B) Proof of Lemma 2:

Let $h_i(n)$, $1 \leq i \leq 4$, $-\infty < n < \infty$ denote the unit pulse response of the i th filter (for causal filters, $0 \leq n \leq \infty$, which falls into the more general case considered here).

Then

$$y_i(n)=\sum_{n'=-\infty}^{\infty} h_i(n-n')u_i(n').$$

i)

$$E\,|y_i(n)|=E\,|\sum_{n'=-\infty}^{\infty} h_i(n-n')u_i(n')|\leqslant \sum_{n'=-\infty}^{\infty} |h_i(n-n')|E\,|u_i(n')|$$

$$\leqslant M_1\sum_{n'=-\infty}^{\infty}|h_i(n')|\leqslant M_1K=K_1<\infty.$$

The last step is obtained since all poles inside the unit circle implies that $\sum_{n'=-\infty}^{\infty}|h_i(n')|\leqslant K<\infty$ for all $1\leqslant i\leqslant 4$ [91].

ii)

$$y_i(n)y_j(m)=\sum_{n'=-\infty}^{\infty} h_i(n-n')u_i(n')\sum_{m'=-\infty}^{\infty} h_j(m-m')u_j(m')$$

$$|E\{y_i(n)y_j(m)\}|=|\sum_{n'=-\infty}^{\infty} h_i(n-n')\sum_{m'=-\infty}^{\infty} h_j(m-m')E\{u_i(n')u_j(m')\}|.$$

By Cauchy-Schwarz inequality $|E\{XY\}|\leqslant(E\{X^2\}E\{Y^2\})^{1/2}$ [92], we have

$$|E\{u_i(n')u_j(m')\}|\leqslant\left|E\{u_i(n')^2\}E\{u_j(m')^2\}\right|^{1/2}\leqslant M_2.$$

Hence

$$|E\{y_i(n)y_j(m)\}|\leqslant M_2|\sum_{n'=-\infty}^{\infty} h_i(n')|\,|\sum_{m'=-\infty}^{\infty} h_j(m')|\leqslant M_2K^2=K_2<\infty.$$

iii) Analogous to i) and ii), we have

$$|E\{y_i(n)y_j(m)y_k(\xi)\}|\leqslant|\sum_{n'=-\infty}^{\infty} h_i(n')|\,|\sum_{m'=-\infty}^{\infty} h_j(m')|\,|\sum_{\xi'=-\infty}^{\infty} h_k(\xi')|$$

$$|E\{u_i(n')u_j(m')u_k(\xi')\}|$$

$$\leq K^3\left|E\{u_i(n')^2u_j(m')^2\}E\{u_k(\xi')^2\}\right|^{1/2} \leq K^3M_2^{1/2}\left|E\{u_i(n')^2u_j(m')^2\}\right|^{1/2}$$

$$\leq K^3M_2^{1/2}\left|E\{u_i(n')^4\}E\{u_j(m')^4\}\right|^{1/4} \leq K^3(M_2M_4)^{1/2}=K_3<\infty$$

iv) Same as before:

$$|E\{y_i(n)y_j(m)y_k(\xi)y_l(\zeta)\}| \leq |\sum_{n'=-\infty}^{\infty} h_i(n')|\,|\sum_{m'=-\infty}^{\infty} h_j(m')|\,|\sum_{\xi'=-\infty}^{\infty} h_k(\xi')|\cdot$$

$$|\sum_{\zeta'=-\infty}^{\infty} h_l(\zeta')|\,|E\{u_i(n')u_j(m')u_k(\xi')u_l(\zeta')\}|$$

$$\leq K^4\left|E\{u_i(n')^2u_j(m')^2\}E\{u_k(\xi')^2u_l(\zeta')^2\}\right|^{1/2}$$

$$\leq K^4\left|E\{u_i(n')^4\}E\{u_j(m')^4\}E\{u_k(\xi')^4\}E\{u_l(\zeta')^4\}\right|^{1/4}$$

$$\leq K^4M_4=K_4<\infty.$$

C) Proof of Lemma 3:

The proof follows Benveniste *et al.* [71] closely.

For simplicity, we drop $\hat{\theta}$ in notations. Using the same definition for matrix norm as before, we have:

i)

$$E\,\|Y(n)Y(n)^T-Y^m(n)Y^m(n)^T\|^2$$

$$=\sum_{i=1}^{n_2}\sum_{i'=1}^{n_2}E[y_1(n-i)y_1(n-i')-y_1^m(n-i)y_1^m(n-i')]^2$$

$$+\sum_{j=0}^{\hat{n}_b}\sum_{j'=0}^{\hat{n}_b} E[y_2(n-j)y_2(n-j')-y_2^m(n-j)y_2^m(n-j')]^2$$

$$+2\sum_{i=1}^{\hat{n}_a}\sum_{j=0}^{\hat{n}_b} E[y_1(n-i)y_2(n-j)-y_1^m(n-i)y_2^m(n-j)]^2$$

$$=\sum_{i=1}^{\hat{n}_a}\sum_{i'=1}^{\hat{n}_a} E[y_1(n-i)y_1(n-i')-y_1(n-i)y_1^m(n-i')+y_1(n-i)y_1^m(n-i')$$

$$-y_1^m(n-i)y_1^m(n-i')]^2$$

$$+\sum_{j=0}^{\hat{n}_b}\sum_{j'=0}^{\hat{n}_b} E[y_2(n-j)y_2(n-j')-y_2(n-j)y_2^m(n-j')+y_2(n-j)y_2^m(n-j')$$

$$-y_2^m(n-j)y_2^m(n-j')]^2$$

$$+2\sum_{i=1}^{\hat{n}_a}\sum_{j=0}^{\hat{n}_b} E[y_1(n-i)y_2(n-j)-y_1(n-i)y_2^m(n-j)+y_1(n-i)y_2^m(n-j)$$

$$-y_1^m(n-i)y_2^m(n-j)]^2$$

$$\leq 2\sum_{i=1}^{\hat{n}_a}\sum_{i'=1}^{\hat{n}_a} E\{y_1(n-i)^2[y_1(n-i')-y_1^m(n-i')]^2+y_1^m(n-i')^2[y_1(n-i)-y_1^m(n-i)]^2\}$$

$$+2\sum_{j=0}^{\hat{n}_b}\sum_{j'=0}^{\hat{n}_b} E\{y_2(n-j)^2[y_2(n-j')-y_2^m(n-j')]^2+y_2^m(n-j')^2[y_2(n-j)-y_2^m(n-j)]^2\}$$

$$+4\sum_{i=1}^{\hat{n}_a}\sum_{j=0}^{\hat{n}_b} E\{y_1(n-i)^2[y_2(n-j)-y_2^m(n-j)]^2+y_2^m(n-j)^2[y_1(n-i)-y_1^m(n-i)]^2\}$$

where the inequality $(a+b)^2 \leq 2(a^2+b^2)$ is again used. Now again using Cauchy-Schwarz inequality as in the proof of Lemma 2, we have:

$$\left\{E \| Y(n)Y(n)^T - Y^m(n)Y^m(n)^T \|^2\right\}^{1/2}$$

$$\leq \left\{ 2 \sum_{i=1}^{\hat{n}_a} \sum_{i'=1}^{\hat{n}_a} \left| E y_1(n-i)^4 \right|^{1/2} \left\{ E\left[y_1(n-i') - y_1^m(n-i') \right]^4 \right\}^{1/2} \right.$$

$$+2 \sum_{i=1}^{\hat{n}_a} \sum_{i'=1}^{\hat{n}_a} \left| E y_1^m(n-i)^4 \right|^{1/2} \left\{ E\left[y_1(n-i') - y_1^m(n-i') \right]^4 \right\}^{1/2}$$

$$+2 \sum_{j=0}^{\hat{n}_b} \sum_{j'=0}^{\hat{n}_b} \left| E y_2(n-j)^4 \right|^{1/2} \left\{ E\left[y_2(n-j') - y_2^m(n-j') \right]^4 \right\}^{1/2}$$

$$+2 \sum_{j=0}^{\hat{n}_b} \sum_{j'=0}^{\hat{n}_b} \left| E y_2^m(n-j)^4 \right|^{1/2} \left\{ E\left[y_2(n-j') - y_2^m(n-j') \right]^4 \right\}^{1/2}$$

$$+4 \sum_{i=1}^{\hat{n}_a} \sum_{j=0}^{\hat{n}_b} \left| E y_1(n-i)^4 \right|^{1/2} \left\{ E\left[y_2(n-j) - y_2^m(n-j) \right]^4 \right\}^{1/2}$$

$$\left. +4 \sum_{i=1}^{\hat{n}_a} \sum_{j=0}^{\hat{n}_b} \left| E y_2^m(n-j)^4 \right|^{1/2} \left\{ E\left[y_1(n-i) - y_1^m(n-i) \right]^4 \right\}^{1/2} \right\}^{1/2}$$

$$\leq \sqrt{2} \sum_{i=1}^{\hat{n}_a} \sum_{i'=1}^{\hat{n}_a} \left\{ \left| E y_1(n-i)^4 \right|^{1/4} + \left| E y_1^m(n-i)^4 \right|^{1/4} \right\} \left\{ E\left[y_1(n-i') - y_1^m(n-i') \right]^4 \right\}^{1/4}$$

$$+\sqrt{2} \sum_{j=0}^{\hat{n}_b} \sum_{j'=0}^{\hat{n}_b} \left\{ \left| E y_2(n-j)^4 \right|^{1/4} + \left| E y_2^m(n-j)^4 \right|^{1/4} \right\} \left\{ E\left[y_2(n-j') - y_2^m(n-j') \right]^4 \right\}^{1/4}$$

$$+2 \sum_{i=1}^{\hat{n}_a} \sum_{j=0}^{\hat{n}_b} \left| E y_1(n-i)^4 \right|^{1/4} \left\{ E\left[y_2(n-j) - y_2^m(n-j) \right]^4 \right\}^{1/4}$$

$$+2 \sum_{i=1}^{\hat{n}_a} \sum_{j=0}^{\hat{n}_b} \left| E y_2^m(n-j)^4 \right|^{1/4} \left\{ E\left[y_1(n-i) - y_1^m(n-i) \right]^4 \right\}^{1/4}.$$

Now from Lemma 2 we know that $E y_2(n-j)^4 < K_4 < \infty$ and

$$Ey_1(n-i)^4=E\left|\sum_{n'=-\infty}^{\infty}h_1(n')x(n-i-n')+\sum_{n'=-\infty}^{\infty}h_2(n')v(n-i-n')\right|^4$$

$$\leqslant E\left|\sum_{n'=-\infty}^{\infty}h_1(n')x(n-i-n')\right|^4+4\left|E\left|\sum_{n'=-\infty}^{\infty}h_1(n')x(n-i-n')\right|^3\left|\sum_{n'=-\infty}^{\infty}h_2(n')v(n-i-n')\right|\right|$$

$$+6\left|E\left|\sum_{n'=-\infty}^{\infty}h_1(n')x(n-i-n')\right|^2\left|\sum_{n'=-\infty}^{\infty}h_2(n')v(n-i-n')\right|^2\right|$$

$$+4\left|E\left|\sum_{n'=-\infty}^{\infty}h_1(n')x(n-i-n')\right|\left|\sum_{n'=-\infty}^{\infty}h_2(n')v(n-i-n')\right|^3\right|+E\left|\sum_{n'=-\infty}^{\infty}h_2(n')v(n-i-n')\right|^4$$

$$\leqslant 16K_4<\infty.$$

Also, by the definition of $y_1^m(n-i)$ and $y_2^m(n-j)$ it is seen that $Ey_1^m(n-i)^4\leqslant Ey_1(n-i)^4\leqslant 16K_4<\infty$ and $Ey_2^m(n-j)^4\leqslant Ey_2(n-j)^4\leqslant K_4<\infty$. Note that K_4 is independent of i, j, and can also be selected so that it is an upper bound for all $\hat{\theta}\epsilon\hat{D}_c$. Let us now examine the following quantity:

$$E[y_1(n-i)-y_1^m(n-i)]^4$$

$$=E\left|\sum_{|n'|\geq m+1}^{\infty}h_1(n')x(n-i-n')+\sum_{|n'|\geq m+1}^{\infty}h_2(n')v(n-i-n')\right|^4$$

$$\leqslant E\left|\sum_{|n'|\geq m+1}^{\infty}h_1(n')x(n-i-n')\right|^4$$

$$+4\left|E\left|\sum_{|n'|\geq m+1}^{\infty}h_1(n')x(n-i-n')\right|^3\left|\sum_{|n'|\geq m+1}^{\infty}h_2(n')v(n-i-n')\right|\right|$$

$$+6\left|E\left|\sum_{|n'|\geq m+1}^{\infty}h_1(n')x(n-i-n')\right|^2\left|\sum_{|n'|\geq m+1}^{\infty}h_2(n')v(n-i-n')\right|^2\right|$$

$$+4\left|E\left[\sum_{n'=m+1}^{\infty} h_1(n')x(n-i-n')\right]\left[\sum_{n'=m+1}^{\infty} h_2(n')v(n-i-n')\right]^3\right|$$

$$+E\left[\sum_{n'=m+1}^{\infty} h_2(n')v(n-i-n')\right]^4$$

$$\leqslant M_4\left\{\left[\sum_{n'=m+1}^{\infty}|h_1(n')|\right]^4+4\left[\sum_{n'=m+1}^{\infty}|h_1(n')|\right]^3\left[\sum_{n'=m+1}^{\infty}|h_2(n')|\right]\right.$$

$$+6\left[\sum_{n'=m+1}^{\infty}|h_1(n')|\right]^2\left[\sum_{n'=m+1}^{\infty}|h_2(n')|\right]^2+4\left[\sum_{n'=m+1}^{\infty}|h_1(n')|\right]\left[\sum_{n'=m+1}^{\infty}|h_2(n')|\right]^3$$

$$\left.+\left[\sum_{n'=m+1}^{\infty}|h_2(n')|\right]^4\right\}$$

For the stable filters considered here, their unit pulse responses always have an exponential form:

$$h_i(n)=\begin{cases}\sum_k C_k p_k^n, & n\geqslant 0, \quad |p_k|<1\\ \sum_l D_l q_l^n, & n<0, \quad |q_l|>1\end{cases}$$

where the summations in k and l are finite. Thus there exists p satisfying $|p_k|\leqslant p<1$ and $|\frac{1}{q_l}|\leqslant p<1$ for all $\hat{\theta}\epsilon\hat{D}_c$ such that $|h_i(n)|\leqslant Cp^{|n|}$. Hence

$$\left\{E[y_1(n-i)-y_1^m(n-i)]^4\right\}^{1/4}\leqslant(M_4)^{1/4}\left\{16\left[\sum_{n'=m+1}^{\infty}Cp^{|n'|}\right]^4\right\}^{1/4}$$

$$\leq 2(M_4)^{1/4} C \sum_{|n'|=m+1}^{\infty} p^{|n'|} = 2C(M_4)^{1/4} \frac{p^{m+1}}{1-p} = \beta_1 p^m$$

for all n, i, and for all $\hat{\theta} \epsilon \hat{D}_c$. By the same reasoning, we have

$$\left\{ E[y_2(n-j) - y_2^m(n-j)]^4 \right\}^{1/4} \leq \beta_2 p^m$$

for all n, i, and for all $\hat{\theta} \epsilon \hat{D}_c$.

Thus, we finally obtain the estimate

$$\left\{ E \| Y(n) Y(n)^T - Y^m(n) Y^m(n)^T \|^2 \right\}^{1/2}$$

$$\leq 4\sqrt{2}(K_4)^{1/4} \sum_{i=1}^{\hat{n}_a} \sum_{i'=1}^{\hat{n}_a} \left\{ E[y_1(n-i') - y_1^m(n-i')]^4 \right\}^{1/4}$$

$$+ 2\sqrt{2}(K_4)^{1/4} \sum_{j=0}^{\hat{n}_b} \sum_{j'=0}^{\hat{n}_b} \left\{ E[y_2(n-j') - y_2^m(n-j')]^4 \right\}^{1/4}$$

$$+ 4(K_4)^{1/4} \sum_{i=1}^{\hat{n}_a} \sum_{j=0}^{\hat{n}_b} \left\{ E[y_2(n-j) - y_2^m(n-j)]^4 \right\}^{1/4}$$

$$+ 2(K_4)^{1/4} \sum_{i=1}^{\hat{n}_a} \sum_{j=0}^{\hat{n}_b} \left\{ E[y_1(n-i) - y_1^m(n-i)]^4 \right\}^{1/4}$$

$$\leq \beta_1 p^m (K_4)^{1/4} [4\sqrt{2}\hat{n}_a^2 + 2\hat{n}_a(\hat{n}_b + 1)] + \beta_2 p^m (K_4)^{1/4} [2\sqrt{2}(\hat{n}_b + 1)^2 + 4\hat{n}_a(\hat{n}_b + 1)] = \alpha_1 p^m$$

for all $\hat{\theta} \epsilon \hat{D}_c$.

ii)

$$\left\{ E \| Y(n) y_1(n) - Y^m(n) y_1^m(n) \|^2 \right\}^{1/2}$$

$$= \left\{ \sum_{i=1}^{\hat{n}_a} E[y_1(n-i)y_1(n) - y_1^m(n-i)y_1^m(n)]^2 + \sum_{j=0}^{\hat{n}_b} E[y_2(n-j)y_1(n) - y_2^m(n-j)y_1^m(n)]^2 \right\}^{1/2}$$

$$= \left\{ \sum_{i=1}^{\hat{n}_a} E[y_1(n-i)y_1(n) - y_1(n-i)y_1^m(n) + y_1(n-i)y_1^m(n) - y_1^m(n-i)y_1^m(n)]^2 \right.$$

$$\left. + \sum_{j=0}^{\hat{n}_b} E[y_2(n-j)y_1(n) - y_2(n-j)y_1^m(n) + y_2(n-j)y_1^m(n) - y_2^m(n-j)y_1^m(n)]^2 \right\}^{1/2}$$

$$\leqslant \left\{ 2\sum_{i=1}^{\hat{n}_a} E\left\{ y_1(n-i)^2[y_1(n) - y_1^m(n)]^2 + y_1^m(n)^2[y_1(n-i) - y_1^m(n-i)]^2 \right\} \right.$$

$$\left. + 2\sum_{j=0}^{\hat{n}_b} E\left\{ y_2(n-j)^2[y_1(n) - y_1^m(n)]^2 + y_1^m(n)^2[y_2(n-j) - y_2^m(n-j)]^2 \right\} \right\}^{1/2}$$

$$\leqslant \sqrt{2}\sum_{i=1}^{\hat{n}_a} \left[Ey_1(n-i)^4 \right]^{1/4} \left\{ E[y_1(n) - y_1^m(n)]^4 \right\}^{1/4}$$

$$+ \sqrt{2}\sum_{i=1}^{\hat{n}_a} \left[Ey_1^m(n)^4 \right]^{1/4} \left\{ E[y_1(n-i) - y_1^m(n-i)]^4 \right\}^{1/4}$$

$$+ \sqrt{2}\sum_{j=0}^{\hat{n}_b} \left[Ey_2(n-j)^4 \right]^{1/4} \left\{ E[y_1(n) - y_1^m(n)]^4 \right\}^{1/4}$$

$$+ \sqrt{2}\sum_{j=0}^{\hat{n}_b} \left[Ey_1^m(n)^4 \right]^{1/4} \left\{ E[y_2(n-j) - y_2^m(n-j)]^4 \right\}^{1/4}.$$

Then by the same reasoning as in i), we immediately have the desired result.

REFERENCES

[1] B. Widrow and M. E. Hoff, Jr., "Adaptive switching circuits," in *1960 IRE WESCON Conv. Rec.*, 1960, pt. 4, pp. 96 - 104.

[2] M. M. Sondhi and D. Mitra, "New results on the performance of a well-known class of adaptive filters," *Proc. IEEE*, vol. 64, no. 11, pp. 1583 - 1597, Nov. 1976.

[3] B. Widrow, "Adaptive filters I: fundamentals," *Rept. SEL-66-126 (TR 6764-6)*, Stanford Electronics Laboratories, Stanford, CA, Dec. 1966.

[4] B. Widrow, "Adaptive filters," in *Aspects of Network and System Theory*, R. Kalman and N. DeClaris, Eds. New York: Holt, Rinehart and Winston, 1971, pp. 563 - 587.

[5] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, "Adaptive antenna systems," *Proc. IEEE*, vol. 55, no. 12, pp. 2143 - 2159, Dec. 1967.

[6] B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, Jr., and R. C. Goodlin, "Adaptive noise canceling: principles and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692 - 1716, Dec. 1975.

[7] D. Morgan and S. Craig, "Real-time adaptive linear prediction using the least mean square gradient algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, no. 6, pp. 494 - 507, Dec. 1976.

[8] J. R. Treichler, "Transient and convergent behavior of the adaptive line enhancer," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, no. 1, pp. 53 - 62, Feb. 1979.

[9] B. Friedlander, "A recursive maximum likelihood algorithm for ARMA line enhancement," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, no. 4, pp. 651 - 657, Aug. 1982.

[10] J. R. Zeidler, E. H. Satorius, D. M. Chabries, and H. T. Wexler, "Adaptive enhancement of multiple sinusoids in uncorrelated noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, no. 3, pp. 240 - 259, June 1978.

[11] L. J. Griffiths, "Rapid measurement of digital instantaneous frequency," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, no. 2, pp. 207 - 222, Apr. 1975.

[12] W. F. Gabriel, "Spectral analysis and adaptive array superresolution techniques," *Proc. IEEE*, vol. 68, no. 6, pp. 654 - 666, June 1980.

[13] M. M. Sondhi, "An adaptive echo canceler," *Bell Syst. Tech. J.*, vol. 46, no. 3, pp. 497 - 511, March 1967.

[14] J. R. Rosenberger and E. J. Thomas, "Performance of an adaptive echo canceler operating in a noisy, linear, time-invariant environment," *Bell Syst. Tech. J.*, vol. 50, no. 3, pp. 785 - 813, March 1971.

[15] D. L. Duttweiler, "A twelve channel digital echo canceler," *IEEE Trans. Comm.*, vol. COM-26, no. 5, pp. 647 - 653, May 1978.

[16] W. B. Mikhoel and F. F. Yassa-Greiss, "A frequency-domain adaptive echo-cancellation algorithm," in *Proc. Intl. Symp. Circuits Syst.*, Chicago, IL, May 1981.

[17] R. D. Gitlin and J. S. Thompson, "A phase adaptive structure for echo cancellation," *IEEE Trans. Comm.*, vol. COM-26, no. 8, pp. 1211 - 1220, Aug. 1978.

[18] R. Wehrmann, J. Van der List, and P. Meissner, "A noise-insensitive compromise gradient method for the adjustment of adaptive echo cancelers," *IEEE Trans. Comm.*, vol. COM-28, no. 5, pp. 753 - 759, May 1980.

[19] R. D. Gitlin and J. S. Thompson, "A new structure for adaptive digital echo cancellation," in *Proc. Natl. Telecomm. Conf.*, Dallas, TX, Dec. 1976.

[20] R. W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, vol. 44, no. 4, pp. 547 - 589, Apr. 1965.

[21] R. W. Lucky, "Techniques for adaptive equalization of digital communication systems," *Bell Syst. Tech. J.*, vol. 45, no. 2, pp. 255 - 297, Feb. 1966.

[22] K. H. Mueller, "A new, fast-converging mean-square algorithm for adaptive equalizers with partial-response signaling," *Bell Syst. Tech. J.*, vol. 54, no. 1, pp. 143 - 153, Jan. 1975.

[23] A. Gersho, "Adaptive equalization of highly dispersive channels for data transmission," *Bell Syst. Tech. J.*, vol. 48, no. 1, pp. 55 - 70, Jan. 1969.

[24] T. Walzman and M. Schwartz, "Automatic equalization using the discrete frequency domain," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 1, pp. 59 - 68, Jan. 1973.

[25] R. D. Gitlin and F. R. Magee, "Self-orthogonalizing adaptive equalization algorithms," *IEEE Trans. Comm.*, vol. COM-25, no. 7, pp. 666 - 672, July 1977.

[26] E. H. Satorius and S. T. Alexander, "Channel equalization using adaptive lattice algorithms," *IEEE Trans. Comm.*, vol. COM-27, no. 6, pp. 899 - 905, June 1979.

[27] F. A. Reed, P. L. Feintuch, and N. J. Bershad, "Time delay estimation using the LMS adaptive filter - static behavior," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 3, pp. 561 - 571, June 1981.

[28] P. L. Feintuch, N. J. Bershad, and F. A. Reed, "Time delay estimation using the LMS adaptive filter - dynamic behavior," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 3, pp. 571 - 576, June 1981.

[29] D. M. Etter and S. D. Stearns, "Adaptive estimation of time delays in sampled data systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 3, pp. 582 - 587, June 1981.

[30] M. Dentino, J. M. McCool, and B. Widrow, "Adaptive filtering in the frequency domain," *Proc. IEEE*, vol. 66, no. 12, pp. 1658 - 1659, Dec. 1978.

[31] N. J. Bershad and P. L. Feintuch, "Analysis of the frequency domain adaptive filter," *Proc. IEEE*, vol. 67, no. 12, pp. 1658 - 1659, Dec. 1979.

[32] E. R. Ferrara, "Fast implementation of LMS adaptive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, no. 4, pp. 474 - 475, Aug. 1980.

[33] S. S. Narayan, A. M. Peterson, and M. J. Narasimha, "Transform domain LMS algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, no. 3, pp. 609 - 615, June 1983.

[34] R. D. Gitlin, J. E. Mazo, and M. G. Taylor, "On the design of gradient algorithms for digitally implemented adaptive filters," *IEEE Trans. Circuit Theory*, vol. CT-20, no. 2, pp. 125 - 136, March 1973.

[35] T. J. Schonfeld and M. Schwartz, "Rapidly converging second-order tracking algorithms for adaptive equalization," *IEEE Trans. Inform. Theory*, vol. IT-17, no. 5, pp. 572 - 579, Sept. 1971.

[36] B. Widrow and J. M. McCool, "A comparison of adaptive algorithms based on the method of steepest descent and random search," *IEEE Trans. Antennas Propagat.*, vol. AP-24, no. 5, pp. 615 - 637, Sept. 1976.

[37] L. J. Grifiths, "A continuously-adaptive filter implemented as a lattice structure," in *Proc. Intl. Conf. Acoust., Speech, Signal Processing*, Hartford, Conn., May 1977, pp. 683 - 686.

[38] C. J. Gibson and S. Haykin, "Learning characteristics of adaptive lattice filtering algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, no. 6, pp. 681 - 691, Dec. 1980.

[39] D. Parikh, N. Ahmed, and S. D. Stearns, "An adaptive lattice algorithm for recursive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, no. 1, pp. 110 - 111, Feb. 1980.

[40] I. L. Ayala, "On a new adaptive lattice algorithm for recursive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, no. 2, pp. 316 - 319, Apr. 1982.

[41] M. L. Honig and D. G. Messerschmitt, *Adaptive Filters: Structures, Algorithms, and Applications*. Boston: Kluwer Academic, 1984.

[42] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1985.

[43] B. Widrow and J. M. McCool, "Comments on 'An adaptive recursive LMS filter'," *Proc. IEEE*, vol. 65, no. 9, pp. 1402 - 1404, Sept. 1977.

[44] P. L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, no. 11, pp. 1622 - 1624, Nov. 1976.

[45] C. R. Johnson, Jr., and M. G. Larimore, "Comments on and additions to 'An adaptive recursive LMS filter'," *Proc. IEEE*, vol. 65, no. 9, pp. 1399 - 1402, Sept. 1977.

[46] D. Parikh and N. Ahmed, "On an adaptive algorithm for IIR filters," *Proc. IEEE*, vol. 66, no. 5, pp. 585 - 588, May 1978.

[47] S. D. Stearns, "Error surfaces of recursive adaptive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 3, pp. 763 - 766, June 1981.

[48] T. Soderstrom, "On the uniqueness of maximum likelihood identification," *Automatica*, vol. 11, no. 2, pp. 193 - 197, 1975.

[49] S. A. White, "An adaptive recursive filter," in *Proc. 9th Asilomar Conf. Circuits, Syst., Computers*, Pacific Grove, CA, Nov. 1975, pp. 21 - 25.

[50] S. D. Stearns, G. R. Elliott, and N. Ahmed, "On adaptive recursive filtering," in *Proc. 10th Asilomar Conf. Circuits, Syst., Computers*, Pacific Grove, CA, Nov. 1976, pp. 5 - 10.

[51] J. M. Mendel, *Discrete Techniques of Parameter Estimation; The Equation Error Formulation*. New York: Marcel Dekker, 1973.

[52] K. J. Astrom and P. Eykhoff, "System identification - A survey," *Automatica*, vol. 7, no. 2, pp. 123 - 162, 1971.

[53] C. R. Johnson, Jr., A. L. Hamm, and J. R. Treichler, "Equation error frequency estimation bias for a white noise obscured sinusoid," in *Proc. 12th Asilomar Conf.*

Circuits, Syst., Computers, Pacific Grove, CA, Nov. 1978, pp. 132 - 136.

[54] C. R. Johnson, Jr., "Adaptive parameter matrix and output vector estimation via an equation error formulation," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 7, pp. 392 - 397, July 1979.

[55] C. R. Johnson, Jr., "The common parameter estimation basis of adaptive filtering, identification, and control," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, no. 4, pp. 587 - 595, Aug. 1982.

[56] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction, and Control*. Englewood Cliffs, NJ: Prentice Hall, 1984.

[57] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. Cambridge, MA: The MIT Press, 1983.

[58] B. Friedlander, "System identification techniques for adaptive signal processing," *Circuits, Syst., Signal Processing*, vol. 1, no. 1, pp. 3 - 41, 1982.

[59] M. G. Larimore, J. R. Treichler, and C. R. Johnson, Jr., "SHARF: An algorithm for adapting IIR digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, no. 4, pp. 428 - 440, Aug. 1980.

[60] C. R. Johnson, Jr., "A convergence proof for a hyperstable adaptive recursive filter," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 6, pp. 745 - 749, Nov. 1978.

[61] C. R. Johnson, Jr., M. G. Larimore, J. R. Treichler, and B. D. O. Anderson, "SHARF convergence properties," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 3, pp. 659 - 670, June 1981.

[62] C. R. Johnson, Jr., and T. Taylor, "CHARF convergence studies," in *Proc. 13th Asilomar Conf. Circuits, Syst., Computers*, Pacific Grove, CA, Nov. 1979, pp. 403 - 407.

[63] D. Parikh, S. C. Sinha, and N. Ahmed, "On a modification of the SHARF algorithm," in *Proc. 22nd Midwest Symp. Circuits Syst.*, Philadelphia, PA, June 1979, pp. 362 - 366.

[64] C. R. Johnson, Jr., "A stable family of adaptive IIR filters," in *Proc. Intl. Conf. Acoust., Speech, Signal Processing*, Denver, CO, 1980, vol. III, pp. 1001 - 1004.

[65] C. R. Johnson, Jr., and T. Taylor, "Failure of a parallel adaptive identifier with adaptive error filtering," *IEEE Trans. Automat. Contr.*, vol. AC-25, no. 6, pp. 1248 - 1250, Dec. 1980.

[66] C. R. Johnson, Jr., I. D. Landau, T. Taylor, and L. Dugard, "On adaptive IIR filters and parallel adaptive identifiers with adaptive error filtering," in *Proc. Intl. Conf. Acoust., Speech, Signal Processing*, Atlanta, GA, 1981, pp. 538 - 541.

[67] C. R. Johnson, Jr., "Adaptive IIR filtering: current results and open issues," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 2, pp. 237 - 250, March 1984.

[68] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-10, no. 4, pp. 461 - 464, Oct. 1965.

[69] P. Stoica and T. Soderstrom, "The Steiglitz-McBride algorithm revisited - Convergence analysis and accuracy aspects," *IEEE Trans. Automat. Contr.*, vol. AC-26, no. 3, June, 1981.

[70] P. Stoica and T. Soderstrom, "Analysis of the Steiglitz-McBride identification method," *Rep. UPTEC 8079R*, Inst. of Tech., Uppsala Univ., Uppsala, Sweden, Sept.

1980.

[71] A. Benveniste, M. Goursat and G. Ruget, "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communication algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-25, no. 6, pp. 1042 - 1058, Dec. 1980.

[72] A. Benveniste and G. Ruget, "A measure of the tracking capability of recursive stochastic algorithms with constant gains," *IEEE Trans. Automat. Contr.*, vol. AC-27, no. 3, pp. 639 - 649, June 1982.

[73] H. J. Kushner and H. Huang, "Asymptotic properties of stochastic approximations with constant coefficients," *SIAM J. Contr. Optim.*, vol. 19, no. 1, pp. 87 - 105, Jan. 1981.

[74] H. J. Kushner and A. Shwartz, "Weak convergence and asymptotic properties of adaptive filters with constant gains," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 2, pp. 177 - 182, March 1984.

[75] H. Fan, "A convergence proof for Fan-Jenkins adaptive IIR filter," submitted to *IEEE Trans. Inform. Theory*, 1985.

[76] H. Fan and W. K. Jenkins, "Adaptive IIR filtering: a new approach," in *Proc. 27th Midwest Symp. Circuits Syst.*, Morgantown, WV, June 1984, pp. 562 - 565.

[77] H. Fan and W. K. Jenkins, "A new adaptive IIR filter," submitted to *IEEE Trans. Circuits Syst.*, 1985.

[78] D. K. Faddeev and V. N. Faddeeva, *Computational Methods of Linear Algebra*. San Francisco: W. H. Freeman, 1963.

[79] P. Billingsley, *Convergence of Probability Measures*. New York: John Wiley, 1968.

[80] R. B. Ash and M. F. Gardner, *Topics in Stochastic Processes*. New York: Academic Press, 1975.

[81] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice Hall, 1980.

[82] C. T. Chen, *Linear System Theory and Design*. New York: Holt, Rinehart, and Winston, CBS College Publishing, 1984.

[83] A. Stefanski and C. N. Weygandt, "Extension of the Steiglitz and McBride identification technique," *IEEE Trans. Automat. Contr.*, vol. AC-16, no. 5, pp. 503 - 504, Oct. 1971.

[84] J. W. Emling and D. Mitchell, "The effects of time delay and echoes on telephone conversations," *Bell Syst. Tech. J.*, vol. 42, no. 6, pp. 2869 - 2891, Nov. 1963.

[85] Technical Staff of Bell Laboratories, *Transmission Systems for Communications*, 5th ed. Bell Telephone Laboratories, 1982.

[86] C. R. Johnson, Jr., and B. D. O. Anderson, "Sufficient excitation and stable reduced-order adaptive IIR filtering," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-29, no. 6, pp. 1212 - 1215, Dec. 1981.

[87] B. D. O. Anderson and C. R. Johnson, Jr., "On reduced-order adaptive output error identification and adaptive IIR filtering," *IEEE Trans. Automat. Contr.*, vol. AC-27, no. 4, pp. 927 - 933, Aug. 1982.

[88] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, no. 8, pp. 1151 - 1162, Aug. 1976.

[89] D. P. Dersvitskii and A. L. Fradkov, "Two models for analyzing the dynamics of adaptation algorithms," *Automat. Remote Contr.*, vol. 35, no. 1, pp. 59 - 67, 1974.

[90] J. J. J. Fuchs, "On the good use of the spectral radius of a matrix," *IEEE Trans. Automat. Contr.*, vol. AC-27, no. 5, pp. 1134 - 1135, Oct. 1982.

[91] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1975.

[92] R. B. Ash, *Real Analysis and Probability*. New York: Academic Press, 1972.

END

DTIC

9-86